walter
sosa escudero

viajar al futuro

(y volver para contarlo)

la ciencia detrás de los pronósticos

siglo veintiuno
editores

walter
sosa escudero

viajar al futuro

(y volver para contarlo)

la ciencia detrás de los pronósticos

siglo veintiuno
editores

Índice

Cubierta

Índice

Portada

Copyright

Este libro (y esta colección)

Dedicatoria

Introducción

1. El futuro llegó hace rato: el azar y los pronósticos

¿Qué es el azar?

El azar más allá de los juegos y la física cuántica

La ley de los grandes números: el teorema fundamental de la ciencia de datos

Big data y el nirvana de la estadística

¿Qué quiere decir “90% de probabilidad de lluvias”?

Maldito Bayes: el portal entre los datos y las probabilidades y los pronósticos

Volver del futuro

2. Encuestas para el pronóstico en tiempos de IA

Encuestas para el pronóstico: ¿una laguna de datos cristalinos?

Encuestas electorales, big data y el tenedor libre de la estadística

¿Tiene sentido hacer encuestas en épocas de IA?

El efecto mariposa de las encuestas mal hechas

Big data persigue al covid-19

Twitter e IA al rescate de las encuestas electorales

Volver del futuro

3. Predicciones para la economía y la guerra

La maldición del Premio Martín Fierro Revelación

La del militar que les grita a los soldados

Regresión a la media y el futuro del capitalismo

La camioneta de Ernesto y los tanques alemanes

Volver del futuro

4. Predecir con probabilidades: pronósticos y pronosticadores

La maquinola de Persi Diaconis

Agitado, no revuelto: el problema de Monty Hall

El efecto Mirtha Legrand y las probabilidades condicionales

La épica librería Daniel

El Robin Hood del barrio de Núñez: azar y talento

Pronósticos versus pronosticadores: “El argentino que predijo la crisis”

Volver del futuro

5. ¿Errar es humano? Errores de pronóstico y el rol de las computadoras

La balada de Laura y Susana: errores de pronóstico

El hombre del rifle: sesgo, varianza y error cuadrático medio

¿Qué se puede hacer, salvo ver películas? Evaluación de errores de pronóstico

¿Es posible evitar la sobreventa en los vuelos?

¡Lo fermizamo’ con alambre! El rol de los algoritmos

Otra vez arroz: el rol de las simulaciones

Revoleando globos terráqueos en Montecarlo

Volver del futuro

6. Complejidad, meteorología y fútbol

Mariposa technicolor: caos, no linealidad y condiciones iniciales

La larga saga de la meteorología

Complejidad y sistemas dinámicos en la predicción del tiempo

Yo era el rey de este lugar: IA para la predicción del tiempo

¡Ghana gana el Mundial! Algoritmos y datos para predecir goles,

gambetas y medallas

Volver del futuro

7. Armas de predicción masiva: los pronósticos y su caja de herramientas técnicas

El campeón de los pronósticos

Regresión: el automóvil de la ciencia de datos

El que mucho abarca, poco aprieta: sobreajuste

La profundidad del ChatGPT: redes neuronales y deep learning

Yo hago ravioles, ella hace ravioles: vecinos cercanos

Volver del futuro

8. Predecir es muy difícil, en particular, el futuro: finanzas, economía y series temporales

Pareidolia y soledad: esos esotéricos gráficos financieros

Series estacionarias y la predecibilidad de lo impredecible

Una alarma en mis oídos: ruido blanco

Random walks, on the rocks

El chofer designado de la Bolsa

ChatGPT para pronosticar la inflación

Volver del futuro

9. La vida es juego: estrategia y predecibilidad

El juego de los penales y las finanzas

El enigma de Alan Turing

La naturaleza imita al arte y la ciencia, a Maradona

Las nubes ignoran a los meteorólogos, pero la economía no ignora a

los economistas

¿Atajar al azar?

Cara de póker y “duro de aleatorizar”

Volver del futuro

10. Manejar mirando el espejo retrovisor: horóscopo, el diario de mañana y cisnes negros

Dios me libre de ser de Libra: la confiabilidad del horóscopo

Kafka llegó hace rato: pronosticar el pasado

Haciendo llover con paraguas

Un cisne negro de 23.388 plumas

¿Cuán normal es la distribución normal?

Volver del futuro

Epílogo. Decálogo del buen pronosticador

Apéndice. Fuentes, lecturas y sugerencias

Agradecimientos

Walter Sosa Escudero

VIAJAR AL FUTURO (Y VOLVER PARA CONTARLO)

La ciencia detrás de los pronósticos

siglo veintiuno editores

Sosa Escudero, Walter

Viajar al futuro (y volver para contarlo)/ Walter Sosa Escudero.- 1ª ed.- Ciudad Autónoma de Buenos Aires: Siglo Veintiuno Editores, 2024.

Libro digital, EPUB.- (Ciencia que Ladra…, serie Mayor / dirigida por Diego Golombek)

Archivo Digital: descarga y online

ISBN 978-987-801-399-2

1. Probabilidad. 2. Cálculos. 3. Estadísticas. I. Título.

CDD 310

© 2024, Siglo Veintiuno Editores Argentina S.A.

< www.sigloxxieditores.com.ar >

Diseño de portada: Pablo Font

Digitalización: Departamento de Producción Editorial de Siglo XXI Editores Argentina

Primera edición en formato digital: noviembre de 2024

Hecho el depósito que marca la ley 11.723

ISBN edición digital (ePub): 978-987-801-399-2

Este libro (y esta colección)

El futuro ya pasó y nosotros ni nos dimos cuenta.

Gianni, protagonista de Nos habíamos amado tanto (Ettore Scola, 1974)

Si juzgamos sabiamente, / daremos lo [aún] no venido /por [ya] pasado.

Jorge Manrique, Coplas a la muerte de su padre

Pronosticar es una ciencia. Pero también un arte. Sí: justo en la intersección entre la economía, las probabilidades y un poquito de juego o de intuiciones, aparecen los pronósticos. Hacer resoluciones de Año Nuevo, por ejemplo, es una especie de pronóstico: prometer ir al gimnasio, bajar de peso, dejar de fumar o ahorrar para comprar la guitarra eléctrica depende de una mirada realista del pasado y del potencial para ajustar las chances que tenemos en el futuro próximo. Pero claro, no se trata solamente de nosotros y nuestras decisiones. Cuando un analista financiero elige un paquete de acciones, o cuando los bancos establecen sus tasas de interés, están jugando con el (nuestro) futuro. Incluso al preparar el presupuesto anual de una empresa o de un país se está pronosticando el futuro de la economía. Desde ya, se asumen posibilidades, inflaciones, tecnologías por venir y otras variables, pero… puede fallar. Podemos equivocarnos en lo que damos por sentado y, sí, pueden aparecer eventos externos que nos tiran el pronóstico… al tacho de los pronósticos errados (que, se sabe, es un tacho muy grande).

Los pronosticadores son científicos, sí, pero también confían en su experiencia y, por qué no, en la intuición para interpretar

correctamente lo que dicen los datos y los modelos predictivos. Algo así hacen los artistas: una vez que manejan expertamente las técnicas clásicas de su arte, se dejan llevar por intuiciones, rompen las reglas, buscan oportunidades. Los humanos tenemos una extraña capacidad de reconocer patrones, aunque sean sutiles, y por ahora somos un poco irreemplazables en esto. Aun en el laboratorio, en condiciones controladas de presión y temperatura (como a la gente de ciencia nos gusta decir), muchas veces las variables harán lo que les venga en gana. Ni que hablar de la vida real… ¿Será que la inteligencia artificial, esa célebre IA, reemplazará los pronósticos basados en la física y, sobre todo, a las personas? En plan de pronosticar sobre ellos, nos atrevemos a anotar que, por un tiempo, es poco probable.

Según el filósofo Karl Popper, la utilidad de una predicción se relaciona con que se pueda refutar. Así, decir que "va a llover en el futuro" es una muy mala predicción, ya que sin duda se va a cumplir, y es irrefutable. En cuanto a las famosas probabilidades de que llueva o no en el futuro cercano, si bien se pueden cumplir o no, resultan mucho más útiles. Sí, como muy bien nos explica Walter Sosa Escudero en las páginas que siguen, saber que hay un 40% de probabilidades de que llueva o truene es tantísimo mejor que no saber nada. Y para saber mejor se necesitan muchos, muchos datos: aquí es donde entra en juego la IA que, en no pocos casos, es el mejor pronosticador de eventos (claro que le faltan el arte o la calle tan necesarios en situaciones como esas y, como dice el autor, "a veces la inteligencia barrial le gana a la artificial").

Cuidado: no es lo mismo predecir (que requiere seguridades que el mundo en general no nos regala, y corresponde al universo de los dioses y los estafadores) que pronosticar, sin tantas certezas y siempre con cierta contaminación de pequeñas y grandes variaciones que juegan con los números a su antojo. Así, el pronosticador juega un juego muy diferente al de los y las profetas y sus profecías: en vez de escudriñar el cielo, los hígados de los carneros o ciertos humos mareadores, busca señales evidentes u ocultas que marquen determinadas tendencias en los resultados. De alguna manera, quien pronostica mapea incertezas, más que verdades absolutas. Como el cono del silencio del Superagente 86 (jóvenes, abstenerse), el/la buen/a pronosticador/a vive en un cono

de incertezas, pero siempre buscando contexto y datos que guíen nuestras intuiciones. Cuanto más podamos reducir nuestro cono, mejores pronósticos podremos hacer. Así, estar atentos a tendencias, crecimientos exponenciales o caídas humillantes nos puede dar un marco para acotar nuestros cálculos para el futuro cercano (y, como aprenderemos en este libro, cuanto más cercano, mejor).

También resulta obvio que las mejores predicciones son locales, y no necesariamente universales; como cuenta el escritor William Gibson, "el futuro ya llegó, solo que aún no está distribuido uniformemente": habrá que estar atentos a esos parches de futuro que van asomando en el clima, los deportes o los mercados para poder entender hacia dónde se van a expandir. Y no solo atender a esos futuros que asoman sino, sobre todo, al pasado. Ya lo dijo el viejo Marshall McLuhan: vamos hacia el futuro mirando por el espejo retrovisor; marchamos al futuro caminando hacia atrás.

Más allá de estas consideraciones, parecería haber gente capaz de pronosticar de manera bastante exitosa. Sin ir más lejos, si estamos viendo una serie, ¿cuántas veces pasó que adivinamos lo que iban a decir los protagonistas? ¿Y qué pasa con la gente que es capaz de "predecir" el costo del petróleo, o el campeón del mundial de jabalina? Así, hay gente que la pega más que la media; son lo que en la jerga suelen llamarse "superpronosticadores": aquellos individuos muy precisos a la hora de visualizar eventos globales en el futuro mediato. Estos superhéroes (y superheroínas) tienden a andar mejor en pruebas de inteligencia, son muy competitivos, están dispuestos a cambiar de opinión si la evidencia lo exige y se entrenan permanentemente para buscar patrones en los datos. Hasta hay cursos para superpronosticar; sin embargo, lo más razonable para entender este fenómeno es comenzar… por este libro, donde Walter nos introduce en la ciencia de los pronósticos, los súper y los no tan súper, los de todos los días y los de las supercomputadoras, los del mapa que nos ayuda a llegar más rápido, los que nos recomiendan llevar un paraguas bajo el brazo y los predictores de goles, gambetas y carambolas.

En definitiva, Walter nos invita a un viaje en el tiempo. Y ya se sabe, según dijeron tanto Eleanor Roosevelt como la tortuga de Kung Fu Panda, "el ayer es historia y el mañana es un misterio,

pero el hoy es un obsequio: por eso se llama presente".

La serie Mayor de Ciencia que Ladra es, al igual que la serie Clásica, una colección de divulgación científica escrita por científicos que creen que ya es hora de asomar la cabeza por fuera del laboratorio y contar las maravillas, grandezas y miserias de la profesión. Porque de eso se trata: de contar, de compartir un saber que, si sigue encerrado, puede volverse inútil.

Ciencia que ladra… no muerde, solo da señales de que cabalga.

Diego Golombek

A Alejandro y Mercedes

Introducción

Salí de Buenos Aires a Tandil a eso de las 7 de la mañana. La idea era evitar el tráfico fuerte del centro de la ciudad y llegar para almorzar. Mientras cargaba nafta cerca de casa recibí un mensaje de mi anfitrión, preguntándome a qué hora llegaría. Puse la dirección de destino en Google Maps, que indicó el horario de las 11.47; en mi respuesta, lo redondeé en un "a eso del mediodía". Subí al auto, crucé el centro de Buenos Aires por la Avenida 9 de Julio, tomé la autopista, seguí la ruta indicada por la app, crucé pueblos, insulté a camioneros que obstruían mi paso, me entretuve con las playlists ruteras que tengo en Spotify, hice una "parada técnica" en alguna estación de servicio, atravesé el centro de Tandil y seguí por un camino rural que a las 11.50 hs –apenas tres minutos más tarde de lo que había predicho la aplicación– me depositó en el coqueto hotel de campo donde tenía que dictar mi conferencia. Buenísimo.

A fines de diciembre de cada año, la revista The Economist publica un informe en el que hace pronósticos en relación con la política, la economía y la sociedad para el año siguiente, titulado "El mundo hacia adelante". En 2022, con respecto a Europa, da detalles del derrotero del Brexit, de Mario Draghi en Italia, del surgimiento de nuevas coaliciones en Alemania. De Putin, solo dice que "renovará sus ataques a las elecciones y a la internet". Y como no lo puedo creer, aprieto "control-F", escribo "Ucrania", y nada: "El diario no hablaba de ti", como dice la canción. Los analistas de The Economist no anticiparon que en cuestión de dos meses –para más precisión, el 24 de febrero de 2022– comenzaría la invasión de Rusia a Ucrania. Malísimo.

Los algoritmos pueden predecir con exactitud cuántos automóviles pasarán por un sensor en una autopista un viernes nublado de marzo a las 13.15, pero nadie tiene la menor idea de cuánto va a valer el dólar de acá a dos semanas. Algún experto la vio venir, pero me gustaría saber cuántos levantaron sus copas a las 12 de la noche

del 31 de diciembre de 2019 y en sus deseos incluyeron “que el covid pase pronto y no haga estragos”. A principios de 2022, los analistas incluían a la Argentina entre los favoritos para ganar el Mundial de Fútbol de Qatar, pero solo luego del agónico penal de Gonzalo Montiel en la muy agónica final los argentinos tuvimos la certidumbre que nuestros corazones buscaban y a la vez se resistían a aceptar. El pronóstico del tiempo predijo correctamente que hoy no iba a llover, pero no tiene la menor idea de qué va a pasar en quince días, para cuando planeo hacer un asadito en casa con unos amigos. El ChatGPT puede lidiar con una infinidad de cuestiones predictivas (incluidos el estado del tiempo y la inflación, como veremos), pero nadie sabe con precisión cuál será el resultado de una elección en unos pocos meses.

Un libro sobre la ciencia de los pronósticos

Propongo en estas páginas una introducción a la ciencia de los pronósticos. En tono coloquial, y con método riguroso, iremos desde cuestiones profundas –en las comarcas de la meteorología o la economía y pasando por los escenarios de la guerra, el futuro del capitalismo o las elecciones– hasta cosas de nuestro mundanal día a día, como pronosticar si alguien va a aprobar un examen o si saldrá cara o ceca al lanzar una moneda.

En tiempos del auge de la ciencia de datos, tiempos veloces y extremos, este libro es una invitación a adoptar una postura inteligente ante la posibilidad de hacer predicciones confiables, sin caer en fanatismos, como el de creer inocentemente que los datos lo podrán todo, o el de quienes aseguran que la inteligencia artificial (en adelante, IA) es nada más que una moda pasajera.

Esta travesía nos permitirá explorar varias cuestiones que atañen a las predicciones. Indagaremos por qué aciertos asombrosos (como los de Google Maps) conviven con yerros u omisiones épicas (como la de Ucrania en The Economist). Veremos que lo que se esconde detrás de las dificultades predictivas es una mezcla de azar, complejidad e interacción estratégica. También nos preguntaremos

qué significa predecir correctamente y –spoiler alert!– veremos que es mucho más que “pegarle” o “acertarle” al futuro, porque las predicciones que importan son las que sirven para tomar decisiones, las que son capaces de provocar cambios de conducta, pero antes de que los eventos sucedan, no con el diario de mañana leído y subrayado.

Esta línea argumental nos llevará a detectar que, en algunas circunstancias, predecir es una cosa y explicar, otra, pero que en muchos contextos son ideas muy relacionadas. Apunto un ejemplo: al ver que la gente sale de sus casas con paraguas podríamos predecir que va a llover, aunque cualquiera sabe que los paraguas no entran en ninguna explicación de por qué llueve. Pero, por otro lado, veremos que la credibilidad de muchas predicciones surge de establecer relaciones (causales) entre distintos fenómenos.

También hablaremos de que hay mucho de mito en las predicciones. Por ejemplo, y contra lo que muchos creen, las predicciones meteorológicas en el corto plazo funcionan muy bien y han mejorado considerablemente en las últimas décadas. Algo similar ocurre con los pronósticos de los resultados en el deporte.

Como es de esperar, dedicaremos varias páginas a discutir el rol que cumplen la estadística, las computadoras, los algoritmos, la IA y la nueva ciencia de datos en los pronósticos. Un tema recurrente es que, aun luego de varios años de avances increíbles en esos campos, muchas cuestiones predictivas permanecen ajenas al escrutinio de los datos y los algoritmos. Un punto central, e igual de recurrente en este libro, es que los principales logros predictivos surgen de una acertada combinación de información, tecnología e inteligencia (tanto de la artificial como de la “natural”, la que provee la ciencia).

Argumentaremos que muchos fenómenos de la sociedad, de los deportes a la economía, son cuestiones esencialmente impredecibles. Ahondar en este tema es delicado, porque pone un freno al sobreentusiasmo que muchos tienen en relación con las capacidades de los datos y la IA.

En este libro, nuestros paseos nos harán conocer además la trastienda de los pronósticos. Indagaremos qué etapas se siguen a la

hora de elaborar predicciones para partidos de fútbol, elecciones, el tiempo, el rumbo del dólar, el precio de las acciones bursátiles y muchos otros aspectos de la vida cotidiana de un país y de su ciudadanía.

Bienvenidos a bordo

En esta travesía que estamos comenzando, activemos el GPS y sumemos algunas aclaraciones. La primera es que no adoptaremos ninguna definición precisa de qué quiere decir "predecir". Una acepción habitual es que predecir tiene que ver con adelantarse al futuro; desde este punto de vista, el paso del tiempo juega un rol obvio. Pero la irrupción de nuevas tecnologías favorece otro significado, tal vez más general: "predecir" es rellenar información faltante, como mencionan Agrawal, Gans y Goldfarb en Máquinas predictivas. Los pronósticos del tiempo (si está por llover o si caerá granizo) entran claramente dentro de la primera acepción. También los de las finanzas, del valor del dólar, de quién ganara las elecciones, etc. La segunda acepción incluye, por ejemplo, los sistemas de recomendaciones, que predicen que a uno le gustará una película sobre la base de las que ya vio y de las que vieron otras personas. También en esta visión caen los sistemas de credit scoring, que pronostican si uno pagará un crédito a la luz de su historia crediticia, sus activos, su educación, el barrio en el que vive, etc. En este libro nos ocuparemos de ambos aspectos de la actividad predictiva.

Hay una tercera acepción, a la cual nos dedicaremos un poco menos, que es el ejercicio contrafáctico del tipo "qué sucedería si…". Por ejemplo, una cosa es predecir cuántos autos pasarán a las 11 de la mañana por el peaje de Hudson en la autopista Buenos Aires-La Plata un viernes nublado de tal mes del año próximo. Y otra, completamente distinta, es predecir cuántos pasarían si se construyese un carril adicional. Este segundo tipo de conjetura, causal, y crucial para la toma de decisiones, cae fuera del ámbito puramente predictivo que atañe a estas páginas.

También es importante aclarar que este es un libro de divulgación científica. Exploraremos cómo pronostican las ciencias, de las más duras, como la física, a las más (injustamente llamadas) "blandas", como la ciencia política. El calificativo "científica" es engañoso y, lamentablemente, se abusa de él para dotar de una pátina de seriedad a actividades sospechosas. Por "científica" entenderemos aquí una forma sistemática, reproducible y transparente de realizar pronósticos, que surgen de datos obtenidos de forma honesta y del uso de métodos claros que pueden ser replicados por cualquier experto de una comunidad amplia. Las corazonadas, los pálpitos, las creencias y otros esoterismos no son parte de este libro, no porque no funcionen, sino porque sus éxitos o fracasos no admiten ni reclaman ninguna explicación.

La IA y su hijo dilecto, el ChatGPT (tomado como genérico representativo de los "grandes modelos de lenguaje"), ocupan un lugar no menor en esta historia. En varias oportunidades discutiremos su enorme potencial y también algunas limitaciones. En lo que respecta a la elaboración de este libro, lo he usado intensamente para varias tareas, que van de las más burocráticas (por ejemplo, todos los gráficos y simulaciones fueron confeccionados por ChatGPT) a otras más profundas, como ayudarme a "discutir" algunas ideas abstractas relacionadas con el ámbito de las predicciones, con resultados dispares pero útiles. Como "ChatGPT" me resulta un término un tanto tosco (amén de condenado a una prematura obsolescencia), lo rebauticé como "Roccatagliata" y así me referiré a él de acá en más. Si hay gente que le pone "García" a su perro o "Alberto" a su hámster, no veo por qué yo no puedo hacer lo propio con mis mascotas computacionales.

Como ha sido una sana práctica en mis otros textos de divulgación, este libro no supone un lector con formación técnica en matemática, probabilidad, computación o estadística. La curiosidad y las ganas de pensar un poquito más allá de lo obvio alcanzan. Ninguno de mis libros anteriores es un requisito para leer este: se los garantizo. Sin embargo, para los que tuvieron la enorme fortuna de toparse con ellos –Qué es (y qué no es) la estadística; Big data, o Borges, big data y yo–, notarán que fue inevitable que algunas historias que involucran al núcleo big data-machine learning-IA reaparezcan,

porque, como dice (¡y repite!) Mirtha Legrand, “hay que repetir, porque el público se renueva”. Aun así, las repeticiones son mínimas.

Más allá de que la bibliografía científica sobre pronósticos es técnicamente ardua, el estilo de este libro es discursivo y coloquial. Así y todo, cada historia y cada ejemplo están escritos con el mayor rigor científico posible. Relegaré al apéndice las referencias bibliográficas que dan sustento a todo el material del libro, así como algunos detalles técnicos para lectores interesados. Como también fue el caso en mis libros previos, mantendré una página con material adicional: lecturas, ejemplos, ejercicios y código computacional. Un gentil QR los llevará a ella. Ostento el sospechoso récord de no haber escrito jamás una nota al pie (ni al final) en ninguno de mis libros, práctica que este mismo no abandonará. ¿Por cábala? Vade retro! Como buen científico, “yo no tengo cábalas, sino costumbres que respeto”, como dijo una vez el entrañable Carlos Salvador Bilardo.

Me pregunto si un domingo de estos Boca derrotará a Estudiantes, cuánto valdrá el dólar de acá a un mes (durante la escritura del libro, estaba a $1025; durante la edición, alcanzó los $1500…), cuándo se van a ir (o volver) los mosquitos y el dengue, y si Trump le ganará a Biden en las elecciones del 5 de noviembre de este año. (Y al leer este párrafo, ustedes se estarán matando de risa, porque tienen el diario de mañana, como Biff en Volver al futuro, ese que regresaba en el tiempo con un librito con los resultados de los partidos a hacerse rico con las apuestas. De hecho, Biden se bajó de la candidatura incluso antes de que estas páginas vayan a imprenta, y con los días empezamos a saber que la nueva candidata era Kamala –y su posible vice, Tim Walz–). Hoy yo solo tengo conjeturas, opiniones, explicaciones, datos y algoritmos, y un pilón de hojas que espero que se conviertan en un libro honesto, uno que los ayude a pensar qué hay detrás de esto de predecir correctamente y qué tiene la ciencia para aportar. Pero certezas, como las de ustedes, no tengo. Los envidio.

1. El futuro llegó hace rato: el azar y los pronósticos

Hoy es 14 de marzo de 2024, y hace cuatro días que llueve en Buenos Aires: más que la Argentina, parece Brasil. Ahora mismo no llueve, pero, así y todo, abro <weather.com> y me dice que las probabilidades de que mañana llueva son del 66%. No dice si va a llover o no, solo habla de probabilidades. ¿Llueve o no llueve? ¿Llevo paraguas o no?

En este primer capítulo nos meteremos con el rol que cumple el azar en la vida cotidiana y, por consiguiente, en los pronósticos. Hay algo contradictorio en intentar predecir un evento atravesado por lo fortuito: si pudiésemos predecirlo exacta y sistemáticamente, muy azaroso no sería. Entonces, todo lo plagado de azar, como el tiempo, la política, la economía, el deporte, ¿es impredecible? Esta es la cuestión que indagaremos en las páginas que siguen.

Comenzaremos explorando qué es esto del azar, tanto en tirar monedas como en la física cuántica y en las cuestiones sociales. Después visitaremos la ley de los grandes números, "el" resultado que garantiza que ciertos eventos fortuitos sean predecibles, y pensaremos qué papel tienen los datos, la estadística, big data y la IA en esta historia. Todo lleva a discutir qué quiere decir una cifra como "un 66% de probabilidades de lluvias" y qué nexo hay entre predecir que un fenómeno suceda y sus chances de ocurrencia; parece lo mismo, pero no lo es.

Entonces, sin más prolegómenos, ahí vamos al desconocido mundo del azar y los pronósticos. Traigan celular, agua potable, un snack y paraguas. Porque de lo único que podemos estar seguros es de que siempre que llovió, llovió.

¿Qué es el azar?

Mucho del material de este libro tiene que ver con el azar, con lo aleatorio, con la suerte, con lo incierto, porque, a la larga, predecir correctamente tiene que ver con lidiar con su presencia. Entender la naturaleza del azar es una empresa intelectual compleja. De hecho, la práctica pedagógica separa dos cosas. Una es qué significan las probabilidades y otra, cuáles son las operaciones que se pueden hacer con ellas. Esta segunda dimensión es matemática y toma forma en el cálculo de probabilidades: el estudio sistemático de todas las operaciones válidas que se pueden realizar, dadas ciertas definiciones y reglas. Por ejemplo, si tiramos dos veces una moneda y el lanzamiento de una es independiente del de la otra, y si las chances de que salga cara en un lanzamiento son del 0,5, las de que salgan dos caras en dos lanzamientos son del 0,25, resultado que viene de multiplicar 0,5 por 0,5. Por simple que parezca, esto surge de aplicar una regla de las probabilidades: la posibilidad de que dos eventos independientes se repitan es el producto de las chances de que cada uno de ellos ocurra. Es la matemática lo que garantiza la validez de este procedimiento (y lo que estudia cualquiera que curse una materia de probabilidad).

Para bien y para mal, el cálculo de probabilidades se abstrae de la primera dimensión: ¿qué significan las probabilidades?, ¿de dónde salen? El razonamiento del párrafo anterior nunca discute de dónde vino que las chances de que al lanzar una moneda salga cara son del 0,5, ni qué significa este número, tampoco cómo se interpreta que las chances de que salga dos veces seguidas cara son del 0,25. La matemática vela por que el razonamiento de que ciertas premisas verdaderas (las chances de que salga cara son del 0,5, los lanzamientos son independientes) a las cuales se les aplica una regla válida (la probabilidad conjunta de que ocurran dos sucesos independientes es el producto de cada una de ellas) sea coherente. La discusión sobre el origen y el significado de las probabilidades excede el marco de la matemática y se acerca, tal vez, al de la

filosofía y de todas las disciplinas en las que el azar tiene un rol, desde la física hasta la economía, pasando por la política, el deporte, la meteorología y la lingüística. Dudo de que haya una rama del conocimiento ajena a lo fortuito.

Algunas ideas básicas sobre el azar nos ayudarán a entender qué incidencia tiene en los pronósticos. Una visión del azar es que no existe. (Imagino perplejidades varias, coronadas con la muy lícita pregunta: "¿Qué?". Tomemos aire. Retomo). Que, en realidad, el azar es una forma simpática de llamar a la ignorancia. Esta es, grosso modo, la interpretación del azar en clave determinista, asociada a Pierre de Laplace, matemático y físico francés, uno de los padres de la probabilidad y la estadística. Desde este punto de vista, decir que las chances de que al lanzar una moneda salga cara son del 0,5 es aceptar que no tenemos un conocimiento preciso de la moneda, de la mesa en la que cae, del movimiento de los dedos que la lanzan, de cómo todo esto interactúa con las leyes de la física, elementos que, con suficiente información, predecirían exactamente el resultado. Entonces, afirmar que al lanzar una moneda interviene el azar es una forma de "tirar la toalla" y decir: "Mirá, es muy complicado predecir qué cara va a salir, porque habría que lidiar con muchos fenómenos difíciles de entender o medir; ergo, vamos a tratar al lanzamiento de una moneda como un fenómeno fortuito". El azar aparece como una representación de ese desconocimiento relacionado con la complejidad. Es el "Dios no juega a los dados" de Einstein: para Dios no hay azar, Dios conoce las leyes de la física porque él las escribió; que nosotros no las sepamos es problema nuestro. Entonces, el azar sería una forma de lidiar con esta ignorancia.

Nombrar a Einstein nos deposita en la otra visión del azar, asociada a la física cuántica. La frase del párrafo anterior la dijo el enorme físico alemán en respuesta a Niels Bohr, quien por ese entonces defendía la visión cuántica de la mecánica, según la cual el azar no es una mera representación, sino que existe como tal; así, ciertos fenómenos están sujetos a un comportamiento esencialmente azaroso sin que por eso exista un mecanismo determinístico detrás: el azar existe per se. "Deje de decirle a Dios qué hacer con los dados", dicen que le retrucó Bohr a Einstein en el caldeado debate. Sí, en la física también se cuecen habas, ¡y qué habas!

El azar más allá de los juegos y la física cuántica

La presencia del azar es algo obvio en juegos, a punto tal que muchos hasta nos hacen el favor de tener la palabra “azar” en su descripción. Y así, puestos entre la espada y la pared, los que nos dedicamos a esto ejemplificamos el azar con juegos de azar, lo que es tan conveniente como redundante (cuando no escapista). Sin embargo, el azar está presente en eventos mucho más pedestres que revolear por los aires una moneda y, aparentemente, menos abstrusos que los que competen a la física cuántica, que espantan a todos salvo a los expertos.

Esta cuestión pide un ejemplo a gritos. Hace más de veinte años que dicto una materia de posgrado que se llama Econometría Avanzada. Si alguien me preguntase “¿qué edad tienen tus alumnos?”, tal vez repreguntaría “¿qué alumnos?”. Si la pregunta se refiere a los que el martes 23 de abril de 2023 estaban presentes en mi clase, la respuesta sería una planilla con las edades de cada uno de ellos. Si, en cambio, la pregunta es “en general”, la respuesta también lo sería.

Supongamos que estoy solo en el aula vacía y que, en términos generales, la pregunta se refiere a predecir la edad del primer alumno que va a llegar al aula (o la del segundo, o del octavo, o de cualquiera, en caso de que los alumnos no sigan un patrón específico para llegar a clase). De ser estrictos, mi respuesta debería ser “no tengo la menor idea”, porque no sé quién es el primer alumno que va a llegar. Pongamos la lupa sobre la frase “no tengo la menor idea”, investiguémosla. En primer lugar, la edad no es un número negativo, así que las edades de 0 para abajo están descartadas, en el sentido de la genial frase “los números negativos son tan poco reales como las vacas negativas”, del no menos genial libro Matemáticas e imaginación, de Edward Kasner y James Newman (si llegan a encontrar la versión prologada por Jorge Luis Borges para su “biblioteca personal”, no duden ni un instante en

comprarla).

Vamos al otro extremo. Según mis fuentes (Roccatagliata, que hace su debut en este libro, ¡aplausos, por favor!), la persona más longeva de la que se tenga registro es una tal Jeanne Calment, que murió en 1997 a la edad de 122 años y 164 días. Con bastante generosidad, digamos entonces que es raro que la edad del primer alumno que entra a mi clase esté por arriba de los 200 años. Sigamos. Es una materia de posgrado, así que resultaría extraño que hubiera chicos menores de 14 años. Además, por mi experiencia, las maestrías técnicas (como las de Economía) se hacen más bien cerca de terminar las carreras de grado, por lo que sería raro ver estudiantes de más de 50 años. Entonces, afinemos más el lápiz: sin mirar planillas de años anteriores, es razonable conjeturar que el grueso de mis alumnos anda entre los 23 y los 30 años. Y en esta disquisición estoy cuando, seis minutos antes de comenzar la clase, entra la primera alumna, Paula Merzaroli, de 25 años.

Volvamos al azar. Fíjense que el largo argumento del párrafo anterior refuta esa primera afirmación de que yo no tenía la menor idea. Por el contrario, sé muchísimo de la edad de mis alumnos, tanto que escribí dos párrafos enteros sobre el tema. Lo que acabo de hacer es exactamente lo mismo que hice cuando describimos las caras de una moneda y las chances que tenían de ocurrir, con más o menos precisión. El azar es mi representación de no saber con precisión la edad de ningún alumno en particular. No hay nada azaroso ni fortuito en la edad de Paula Merzaroli desde la perspectiva de Paula Merzaroli, sino que el azar es una forma de indicar, desde mi punto de observación, que sé mucho (pero no lo suficiente) acerca de su edad. Podría haber azar desde la perspectiva de Paula, por ejemplo, si ella fuese adoptada y sus padres no conociesen con precisión su fecha de nacimiento, situación recurrente en familias de padres adoptantes, como la mía. Más allá de esta circunstancia, en muchos fenómenos sociales el azar es una forma de representar aquello que es incierto, sin importar si el azar realmente existe (como afirma la física cuántica) o si es una construcción útil para referirse a lo impreciso (como sostiene la visión determinista).

¿Y qué tiene que ver esta discusión cuasifilosófica con el hecho de

pronosticar? Todo. A la larga, la forma de lidiar con lo incierto es asignarle probabilidades. La actividad de acertar las chances cae en el corazón de la práctica predictiva, y es una parte importantísima de la tarea que hace Netflix cuando recomienda películas (relacionadas con las chances de que a uno le guste o no una película), un banco cuando decide si otorgar un crédito (según las probabilidades de que quien lo solicita lo pague), o el Servicio Meteorológico Nacional cuando emite un alerta por una tormenta severa sobre la base de las posibilidades de que se desencadene.

Entonces, una cuestión clave es cómo se calculan las chances de que algo ocurra. Ha llegado el momento: nos toca hablar de la "llave maestra" de la ciencia de datos y de la estadística, la que relaciona el mundo real (de los datos) con el abstracto (de las chances y la matemática), la que permite calcular probabilidades de lluvia y, con alguna astucia, predecir si efectivamente va a llover.

La ley de los grandes números: el teorema fundamental de la ciencia de datos

Busco en Google “teorema fundamental de”, y la lista que salta es larga: del cálculo, del álgebra, de la hidrostática, de la aritmética, etc. Es como que cada rama de la matemática tiene su resultado basal, como cada país su Padre de la Patria (¿habrá madres?).

Y aquí es donde hacen su aparición estelar la estadística y la ciencia de datos. Desde cierto punto de vista, “ciencia de datos” es nada más que una nueva denominación para la estadística, disciplina crucial que viene haciendo de las suyas desde hace siglos. “Ciencia de datos es estadística con masa madre”, dijo alguien en las redes sociales. Desde una perspectiva más realista, la nueva denominación se refiere a que “la cuestión de los datos” es ahora una disciplina integradora, en la que intervienen la computación, la matemática, el diseño, la comunicación y el conocimiento específico (de la sociología a la física, de la lingüística a la biología). Es desde esta perspectiva interdisciplinar que la ciencia de datos tiene algo para aportar más allá de la estadística. De aquí en más, diremos “estadística” y “ciencia de datos” como si fuesen sinónimos, con las aclaraciones del caso solo cuando sean inevitables.

En todo caso, el negocio de estas disciplinas tiene que ver con aprender de los datos, con que de observar un fenómeno repetidas veces se pueda inferir algo. Y si en la sección anterior discutimos sobre el rol trascendental de las probabilidades en la tarea predictiva, la promesa más relevante de la ciencia de datos se relaciona con la posibilidad de aprender probabilidades a partir de observar la realidad.

Vayamos una vez más al ejemplo azaroso por excelencia. Supongamos que queremos saber cuáles son las chances de que salga cara al tirar una moneda. En estadística y probabilidad hay dos rutas para aproximarse a este problema. Una es el conocimiento experto. Podríamos pensar en un conciliábulo de sabios de la física,

la matemática y la anatomía, a quienes se les provee la moneda para que la estudien, pero que no pueden acudir a revolearla. Luego de una larguísima disquisición, muy posiblemente estos sabiondos concluyan que, sobre la base de todo lo que conocen de sus disciplinas, no hay razón para argumentar que una cara es distinta de la otra, ni que una persona normal pueda controlar sus dedos para favorecer una cara, ni que la física del entorno haga que un resultado tenga más chances de ocurrir que el otro. De tener que apostar a cara o ceca, se declararían indiferentes.

Ojo, indiferentes no es ignorantes: llegar a la conclusión de que no hay más chances de que salga cara que ceca es un enorme paso adelante. En síntesis, esta ruta desembocaría en la confirmación de que las chances de que salga cara son del 0,5, porque así deberían ser también las de que salga ceca. Es una prueba por autoridad. Argumentativa, explícita, científica, pero por autoridad. Se basa en análisis, en introspección, en discusión. Es discrecional pero no arbitraria. A falta de mejor nombre, a este tipo de fundamentación se la llama subjetiva, no porque haya caprichos involucrados sino porque surge de la opinión de alguien.

Pero hay otra forma de afrontar esta cuestión: lanzando la moneda. Podríamos revolear la moneda repetidas veces y anotar si sale cara o ceca. Y cabe esperar que, de continuar ad nauseam con este procedimiento, saldrá ceca el 50% de las veces. Pueden ponerse a tirar una moneda y hacer el experimento. Acá lo hizo Roccatagliata 500 veces (le puso 1 a cara y 0 a ceca):

10100101111110100011100101011010110010110010111100000001011(

Si sacan el promedio, da que el 50,5% de las veces salió cara. Si se sienten aventureros, pueden repetir el experimento con las variantes que prefieran:

hacerlo muchas veces, siempre con 500 lanzamientos,

achicar y agrandar la cantidad de lanzamientos,

hacerlo muchas veces aumentando progresivamente la cantidad de lanzamientos,

y fijarse qué pasa.

La ley de los grandes números (LGN) es uno de los pilares de la probabilidad; en el contexto de nuestro ejemplito, dice que si ustedes pudiesen lanzar la moneda tal como lo hicimos (con la mano o con la compu), pero infinitas veces, el promedio de los resultados debería coincidir con las chances de que salga cara. Lo que hemos hecho, asistidos por la LGN, es averiguar la probabilidad al repetir un experimento. A este enfoque se lo llama objetivo, porque no interviene la opinión de nadie. Es más, el resultado no debería cambiar si la moneda es lanzada por cualquiera de ustedes o un matemático profesional (¡nuevo spoiler alert!: más adelante los sorprenderé con que esto no es taaaaaan así; nunca jueguen al azar contra un matemático profesional).

Objetivo y subjetivo es el "Boca versus River" de la estadística. De versar este libro sobre probabilidades, esta discusión ocuparía un lugar central. Más allá de lo semántico, no resulta obvio que objetivo es mejor que subjetivo. En el primer enfoque, no interviene la opinión de nadie, pero hay que tirar la moneda infinitas veces. En el subjetivo, hay que analizar, pero no hace falta tirar la moneda. Lo importante para la práctica es que la LGN es el "portal" que une el mundo de lo abstracto (las probabilidades) con el de la realidad (los datos) y con la frecuencia relativa o proporción de veces que ocurre un fenómeno. La probabilidad es un número abstracto que refiere a una característica de un evento. La frecuencia relativa surge de los datos, de ver la realidad replicarse.

Un pequeño punto técnico: la LGN tal cual la discutimos no es el teorema fundamental de la estadística. En realidad, semejante honor le corresponde al teorema de Glivenko-Cantelli, una versión matemáticamente más precisa de la LGN.

Big data y el nirvana de la estadística

Atravesar esta puerta mágica con nuestra “llave maestra” no es gratuito. En primer lugar, en su versión estricta, la LGN requiere que la moneda se lance infinitas veces. (De hecho, antes tiramos la moneda 500 veces y la frecuencia no dio exactamente 0,5 sino 0,505). La relevancia de la LGN es que garantiza que la discrepancia entre la frecuencia relativa es insignificante cuando se lanza la moneda muchas veces. En otras palabras, en la realidad, cuando uno no tiró infinitas veces la moneda, la LGN provee una aproximación útil a la verdadera probabilidad.

Este es el auténtico partido que juega la ciencia de datos: dar aproximaciones útiles aun cuando no sean estrictamente verdaderas. La estadística no pretende ser verdadera sino útil. Una encuesta no reemplaza a la población; tan solo la aproxima, en el sentido de la frase del estadístico George Box, que quienes nos dedicamos a los datos tenemos tatuada en el cuerpo o en la mente: “Todos los modelos están mal, pero algunos son útiles”.

El problema de atrapar lo infinito con una cantidad finita de datos parece tener los días contados en tiempos de big data, el fenómeno de datos masivos producto de interactuar con dispositivos interconectados (sensores, redes sociales, etc.) En efecto, muchos problemas viejos de la estadística fueron resueltos por esta revolución de datos masivos. Aun así, lamentablemente, la LGN viene con “letra chica”, de esa que pocos leen y que pone límites al entusiasmo de big data.

En su versión más simple, la LGN demanda dos condiciones técnicas: que los lanzamientos sean

idénticamente distribuidos,

independientemente distribuidos.

En criollo, y en el contexto del lanzamiento de la moneda, estas condiciones implican que cada lanzamiento ocurra en el mismo contexto (misma moneda, mismo lanzador, etc.) y que no implique un aprendizaje o, de forma similar, que cada lanzamiento sea independiente de cualquier otro, cosa que parece razonable en el caso de las monedas.

Big data tiende a darse de bruces contra estos requisitos. Su naturaleza anárquica y espontánea hace que muchas veces los datos sean muy dependientes y heterogéneos: la posibilidad de aprender a partir de ellos se relaciona con que el mismo fenómeno se repita muchas veces y de manera independiente. Como apuntábamos, la LGN es la llave maestra de ese aprendizaje, es lo que permite estimar probabilidades de observar la realidad y de ahí su rol fundamental en la ciencia de datos. Cuando big data se aviene a los requisitos de la LGN, “big” significa, realmente, mucha información. Cuando no, es solo un montón de datos.

¿Qué quiere decir “90% de probabilidad de lluvias”?

Dice el pronóstico: “90% de probabilidad de lluvias, vientos del Sur a 62 km/h, poco cambio de temperatura”. OK, pero ¿va a llover o no? Una cosa es que un evento ocurra y otra las chances de que ocurra. Se habrán dado cuenta de que, por ejemplo, el Servicio Meteorológico Nacional habla de las chances de lluvia, no de si va a llover o no. Desde la perspectiva binaria que muchos tenemos sobre la vida y el funcionamiento del mundo, reducir el pronóstico de lluvia a sus probabilidades suena a timorato, a dubitativo.

Y algo de eso hay, pero ordenar esa distinción es un paso importantísimo para comprender qué significa pronosticar correctamente y cómo se diseñan y evalúan los pronósticos. La característica esencial de un evento azaroso es que es imposible predecir con exactitud qué ocurrirá. Es un poco por contradicción: si fuese completamente predecible, no sería azaroso, qué tanto.

Entonces, cuando alguien dice que hay 90% de probabilidades de que llueva, ¿qué cuernos quiere decir? La discusión anterior sobre el origen de las probabilidades sugiere dos alternativas para pensarlo. Una es que, como con la moneda, del análisis de los factores que influyen en el tiempo se concluye que, de apostar, conviene hacerlo con 90 fichas a que llueve y con 10 a que no. Esta sería la interpretación subjetiva de la probabilidad. La otra es empírica. Los meteorólogos usan complicadísimos modelos para representar el tiempo y luego los ponen a operar (simular), de forma conceptualmente idéntica al lanzamiento de las monedas. Entonces, “90% de probabilidad de lluvias” quiere decir, poco más o menos, que en el 90% de las simulaciones en las que las condiciones del tiempo eran las del día de hoy, llovió.

Ahora bien, ¿por qué no se juega el Servicio Meteorológico Nacional? Bueno, sí se juega, pero no el todo por el todo, a decir si va a llover o no, sino a que las chances de lluvia que publica estén bien calculadas. En el mismo sentido en que, poco antes, cuando

examinamos una moneda con detalle, nos jugamos a decir que no había más chances de que saliera cara que ceca, aun cuando, una vez que lanzamos la moneda, o sale cara o sale ceca.

Que eso suceda no refuta que las chances de que saliera cara o ceca fueran iguales. Esta es una confusión recurrente, que noté y oí aun en gente muy ilustrada. Y que muchos caigan en ella se debe a que la cuestión es difícil. Digámoslo así: una cosa es el rifle y otra, los disparos. Las probabilidades son asignables al rifle, no a los disparos. Cuando aseveramos que la probabilidad de que, al tirar una moneda, las chances de que salga cara son del 0,5, esa aseveración atañe a una característica del hecho de lanzar una moneda, y no a algún lanzamiento en particular.

Decir que las chances de que al lanzar una moneda salga cara son del 0,5 es muchísimo más que decir “no tengo idea”. Acudo a otro ejemplo. Si en el centro de la ciudad de Buenos Aires le preguntáramos a una persona al azar si está desempleada, las chances de que lo esté no son del 50%. Sabemos eso porque la tasa de desempleo en la ciudad es de aproximadamente el 7,5% de la población. En otras palabras, si bien no sabemos con exactitud si una persona que deambula por Buenos Aires está desempleada o no, sí sabemos que es mucho más probable que no lo esté. Y al respecto no podríamos decir que no tenemos “la menor idea”.

Este es tal vez el error más frecuente en relación con el azar y los pronósticos: creer que porque no podemos predecir exactamente un fenómeno tampoco podemos saber cosa alguna de él. De hecho, las probabilidades son una forma de aproximar y mejorar este conocimiento.

Maldito Bayes: el portal entre los datos y las probabilidades y los pronósticos

Para colmo, hay un nexo muy fuerte entre predecir un evento y sus chances de ocurrencia. La negociación entre estas dos magnitudes nos lleva al corazón mismo de la IA y del machine learning. A modo de ejemplo, veamos un mecanismo de credit scoring, de esos que usan los bancos para saber si un cliente va a pagar un crédito: tiene como input información del cliente (sus activos, su educación, su historia crediticia, etc.) y como output una respuesta binaria (va a pagar o no va a pagar el crédito). Dicho así, parece que, a diferencia del Servicio Meteorológico Nacional, este algoritmo se la juega, porque dice "paga o no paga" y no se anda con vueltas.

Para los propósitos de este libro, es crucial entender el nexo entre predecir un evento y las chances de que ocurra. Les propongo un ejercicio mental. Supongamos que alguien nos contrata para predecir si en determinado día va a llover y que sabemos fehacientemente que las chances de que eso ocurra son del 80%. En este contexto, hay cuatro posibilidades:

pronosticamos lluvia y llueve,

pronosticamos lluvia y no llueve,

pronosticamos que no llueve y llueve,

pronosticamos que no llueve y no llueve.

Como verán, en los escenarios 1 y 4 predecimos correctamente, mientras que en 2 y 3 nos equivocamos.

Supongamos que quien nos contrata nos paga $100 si acertamos, y nada si no. Entonces, tenemos que decidir si nos jugamos por decir si llueve o no. Es crucial ver que una vez que nos jugamos por un pronóstico, de los cuatro escenarios posibles pueden aparecer solo dos. Por ejemplo, si pronosticamos lluvia, pueden ocurrir 1 (¡acertamos!) o 2 (siamo fuori della coppa).

El clasificador de Bayes (técnica que se basa en las investigaciones de Thomas Bayes, religioso y matemático británico, y figura fundamental de la historia de la probabilidad) dice que, bajo algunas condiciones razonables, conviene predecir que va a ocurrir el evento más probable. Por ejemplo, si la probabilidad de que llueva es del 90% y la de que no llueva es de un 10%, entonces conviene jugarse a predecir lluvia. Exactamente lo mismo ocurre si la chance de lluvia es de un 51%. Esta regla puede parecer trivial, pero no es tan así. De hecho, el razonamiento que la sostiene descansa en algunos supuestos delicados.

El valor esperado de un suceso aleatorio es lo que se espera ganar. Volvamos al ejemplo de las monedas: si al tirar una, ganamos $100 en caso de que salga cara y $0 si sale ceca, el valor esperado del juego es $50. Hay varias formas de entender de dónde surge esa cantidad; una, simple, es apelar a la ley de los grandes números: de jugar repetidas veces, la mitad de las veces ganaríamos $100 y la otra, nada; así, en promedio, ganaríamos $50.

El clasificador de Bayes dice que, con algunos requisitos técnicos, predecir el evento más probable produce el valor esperado más alto. Parece un tecnicismo, pero la regla de Bayes explica la relación que hay entre predecir eventos y sus chances de ocurrencia. De hecho, cuando se sigue a rajatabla esta regla, predecir las chances de que algo ocurra depende casi exclusivamente de predecir sus probabilidades.

Y la pregunta de por qué el Servicio Meteorológico Nacional no va de las probabilidades a los eventos se relaciona en parte con lo que

discutimos; en realidad, con su letra chica. Implícita en la regla de Bayes hay una forma de evaluar los aciertos y los errores. Por ejemplo, para que la regla opere es necesario suponer que los aciertos y yerros importan simétricamente, es decir, que errar por exceso molesta lo mismo que por defecto.

Como ampliaremos más adelante, en muchas circunstancias la evaluación de yerros o aciertos no necesariamente es simétrica. Por ejemplo, llegar diez minutos tarde puede no ser lo mismo que llegar diez minutos temprano. Más aún, sin pruebas concluyentes, al derecho no le da igual encerrar a un inocente que dejar suelto a un potencial culpable. Otro tanto sucede con la apelación al valor esperado. Que el criterio de Bayes descanse en él presupone cierta conducta orientada hacia el riesgo, situación harto estudiada en economía, psicología y las disciplinas del comportamiento. Así, con tal de enfrentarse a un evento azaroso, alguien con fuertes preferencias por el riesgo aceptaría una menor remuneración esperada, como quien paga una entrada al casino sabiendo que el valor esperado de jugar a la ruleta es 0. Asimismo, la regla de Bayes presupone cierta racionalidad en quien decide, por lo menos la que hace falta para involucrarse en sacar las cuentas (o proceder como si las hiciese) de qué se gana y qué se pierde.

En definitiva, el uso del clasificador de Bayes da mucho por sentado acerca de como se evalúan los pronósticos. De ahí que pasar de decir “90% de probabilidades de lluvia” a decir “va a llover” implica dar un paso enorme, que va mucho más allá de lo que hace falta para medir las probabilidades. Por ejemplo, si aun con 90% de probabilidad de lluvia alguien se pusiera a llorar por tener que cancelar un asado, le pronosticaríamos “no llueve” y, con las mismas chances, a alguien como mi esposa, que se espanta de que le caiga apenas una gotita, le diríamos “llueve”. Entonces, para las mismas probabilidades estamos generando dos pronósticos muy distintos. En este contexto, tal vez lo más razonable sea difundir las probabilidades y que cada uno decida su propio pronóstico y actúe en consecuencia.

Así, la negativa a dar el salto de calcular probabilidades a predecir eventos es un principio de honestidad intelectual, no de cobardía. Cuando ante el lanzamiento de una moneda un científico dice “50%

de que salga cara" y no se la juega por cara o ceca, no es timorato sino sincero en relación con lo que está a su alcance: mecanismos, razonamientos, datos y algoritmos que lo llevaron a calcular una probabilidad. Saltar de las chances de que salga cara a decir "cara" entraña meterse con los costos y los premios relativos de las consecuencias de predecir un resultado o el otro, los cuales, como vimos, no necesariamente están bajo control del investigador, sino que muchas veces dependen de las preferencias de quien juzga y usa el pronóstico. En el capítulo 5 analizaremos que la evaluación de los pronósticos no es independiente de las preferencias de quien los juzga, y que, en la práctica, se "tunean" los algoritmos predictivos para sopesar los costos y beneficios de las predicciones. Adaptar los mecanismos predictivos a estas cuestiones relacionadas con los costos y los beneficios de errar y acertar es una parte crucial de la actividad predictiva.

Volver del futuro

“Datos, no anécdotas”, repiten muchos, relativizando el rol de la observación. La tarea de la estadística (y, en estos tiempos, de big data y de la IA) es transformar una montaña de anécdotas en información útil. En este capítulo hemos revisado las siguientes ideas:

El azar es una convención (enfoque determinista) o una realidad (enfoque propio de la física cuántica) y su presencia es uno de los factores clave de las dificultades a la hora de pensar la posibilidad de hacer predicciones confiables.

Las probabilidades son el esqueleto de lo incierto. Conocer las probabilidades de que un evento ocurra implica entrar en el corazón de la actividad predictiva.

La ley de los grandes números (LGN) es la llave maestra que conecta el mundo real (de los datos) con el de las probabilidades. Es la idea que permite aprender chances a partir de ver que un evento se replica.

Big data no necesariamente es el fin de la estadística ni nos deposita en el “nirvana de los datos”, porque de por sí big data no es todos los datos. Es una colección enorme de datos anárquicos que no necesariamente satisfacen las condiciones bajo las cuales opera la ley de los grandes números.

Así, saber aprovechar las ventajas de la masividad de datos, sin que aparezcan sus dificultades, es el gran desafío de la ciencia de datos para la cuestión predictiva.

En lo que respecta a los pronósticos, hay un nexo directo entre predecir un evento y predecir las probabilidades de que el evento ocurra. El clasificador de Bayes es uno de estos vínculos: de acuerdo con esta regla, se trata de predecir el evento con mayores chances de ocurrir.

Desde esta perspectiva, predecir implica antes haber predicho probabilidades.

2. Encuestas para el pronóstico en tiempos de IA

En 2012, el talentoso músico argentino Pedro Aznar realizó un merecido homenaje al icónico Luis Alberto Spinetta, en el que reinterpretó parte de su vasto repertorio. Como todos saben, la obra de Spinetta es compleja, desde todo punto de vista. Aznar abrió su concierto con “Cantata de puentes amarillos”, tal vez la más difícil y profunda de las composiciones del “Flaco”: más de nueve minutos de duración, más que abundante en cambios de estructura, con una lírica refinadamente sinuosa y una armonía desafiante. Dice Aznar que comenzó por la más brava para liberarse rápido de la presión emocional de semejante patriada: después de la “Cantata”, todo debería ser más fluido.

En esa línea, nuestro recorrido por el mundo de los pronósticos comienza “por la más difícil”: el farragoso terreno de las encuestas, en particular, las relacionadas con los procesos electorales. En el capítulo anterior examinamos que lo que subyace a la naturaleza impredecible de muchos fenómenos es el azar. Si el azar está presente en las cuestiones de la atmósfera, cuando hablamos de la política y las cuestiones sociales en general, hay para hacer dulce y repartir entre los vecinos.

Este capítulo se propone como una introducción al funcionamiento de las encuestas científicas: en especial, veremos en qué condiciones es posible esperar resultados confiables de ellas. Quizá “la” pregunta sea si, en pleno auge de redes sociales e IA, las encuestas tal como las conocemos tienen los días contados. Veremos que no necesariamente.

Encuestas para el pronóstico: ¿una laguna de datos cristalinos?

¿Tienen alguna idea de con cuántas observaciones se hace una encuesta electoral profesional, de esas que invaden los medios antes de cualquier elección? ¿Y de cuántos hogares son encuestados para medir la pobreza? ¿Y de cuánta sangre le sacan a un adulto sano para un análisis rutinario? Les cuento los números. En la Argentina, una encuesta de opinión publica se hace con unos 1000 casos (sí, mil). La pobreza en el Área Metropolitana de Buenos Aires se mide con una muestra de 3657 hogares, cuando en la región hay aproximadamente 4.000.000. Y para un hemograma estándar te sacan unos 5 ml de sangre. Siempre que empiezo mis charlas y seminarios con estas preguntas y estas precisiones, oigo fuerte el ruido de las mandíbulas dando contra el piso y flota la impresión de que es poco, poquísimo.

En cuestiones sociales, se apela a encuestas porque acceder a la población es costosísimo; por eso mismo, los censos –que demandan parar a un país entero– se realizan solo cada diez años. Entonces, la encuesta demuestra ser una estrategia inevitable, que intenta ir de la parte al todo. En este contexto, hablar de poco o mucho es necesariamente una cuestión de grados, de cuán dispuestos estamos a tolerar una “mentira útil y barata” (la que provee la encuesta) en comparación con la imposibilidad de acceder a la población.

Es interesante que la LGN, que vimos en el capítulo anterior, es algo así como un dímer, perilla que nos permite negociar cuánto del aumento del tamaño de la muestra se traduce en mejores predicciones y estimaciones. Con la perilla en mínimo, no hay datos, no cuesta nada, pero tampoco hay información; si la llevamos al máximo, tenemos todos los datos, toda la información, pero el costo puede ser exorbitante. Es en algún lugar intermedio donde toleramos el error implícito en una encuesta si permite aprender algo con un ahorro considerable de recursos. La encuesta honesta, científica, es una reacción pragmática ante la imposibilidad de

acceder a todos los datos.

A modo de ejemplo, la Encuesta de Satisfacción Política y Opinión Pública (Espop) “versa sobre la satisfacción con la marcha general de las cosas, el desempeño de los poderes políticos y las políticas públicas, la aprobación del gobierno y sobre la opinión respecto de los principales líderes políticos nacionales”, como indica su sitio web. Es elaborada periódicamente por un equipo de la Universidad de San Andrés, bajo la dirección del prestigioso politólogo Diego Reynoso, una de las máximas autoridades en cuestión de opinión pública. Cada vez que se publica, la agenda de Reynoso explota de entrevistas y notas, máxime en épocas electorales. La encuesta se realiza con tan solo 1107 casos, gracias a un puntilloso protocolo de recolección de datos que también puede encontrarse en la página de la encuesta (dejo el link en el apéndice).

En relación con las elecciones argentinas de 2023, con el “diario de ayer” (los datos se recabaron entre el 3 y el 8 de noviembre de ese año) la encuesta de Espop decía que “para el ballottage del 19 de noviembre, teniendo en cuenta las respuestas crudas, el 37% de los encuestados manifestó que va a votar por Javier Milei, mientras que un 31% va a votar por Sergio Massa”. El 19 de noviembre apareció el “diario de mañana” y pasó lo que pasó. La encuesta no ve a indecisos y otros no votantes, pero cualitativa y cuantitativamente sugiere, predice y adelanta el resultado; todo, con solo 1107 datos.

La pregunta obvia es si “esto de las encuestas” no es un artefacto social condenado a una prematura obsolescencia, algo más propio de épocas sin computadora, sin internet, sin IA, resabio de situaciones de intensa escasez de datos. Les pido que, por unas páginas, contengan (contengamos) el “obvio” que podríamos dar por respuesta; después retomamos la discusión. Mientras, propongo que recuerden que algo parecido decíamos todos en relación con los vinilos cuando cerca de 1990 se popularizaron los CD.

Encuestas electorales, big data y el tenedor libre de la estadística

“¡Ey, ma! ¡Mirame! ¡Voy camino a la tierra prometida!”, cantaba Bon Scott, el carismático vocalista de AC/DC, en “Highway to Hell”. “¡Se nota!”, contestaron más tarde sus críticos conservadores, espantados por el satanismo. Scott falleció siete meses después del lanzamiento del épico álbum homónimo de la banda de los hermanos Young: se había distraído con unas copas en su camino… quién sabe hacia dónde.

Lo mismo, pero en sentido literal, deberían cantar los estadísticos ante el aluvión de big data. Como mencionamos, históricamente la estadística operó sin salir de una suerte de condena infernal a vivir con datos limitados. Por eso, en la tierra prometida de la estadística están todos los datos.

Una lectura inocente del fenómeno de big data es que nos ubica a solo un par de estaciones del Edén de la información. Ya subidos al trencito de la alegría de la IA y los algoritmos, Viktor Mayer-Schönberger y Kenneth Cukier escribieron que “hacer una muestra aleatoria en la era de big data es como aferrarse a un caballo en la era del automóvil” en su premonitorio (y algo oportunista) libro sobre la cuestión. Y para reforzar el punto, siguieron: “El concepto de muestreo ya no tiene sentido en estas épocas en que podemos apelar a una masiva cantidad de datos”.

Y acá quien sonríe socarronamente es Xiao-Li Meng, director del Departamento de Estadística de la Universidad de Harvard, mientras ve cómo, respecto de las encuestas online (en particular, las electorales), el tren de big data va derechito a juntarse con Bon Scott, a menos que haga algo.

En 2018, Meng, un enorme intelectual (y fanático de nuestro Jorge Luis Borges), tiró la bomba atómica de la ciencia de datos, en un trabajo cuyo título podemos traducir como “Paraísos estadísticos y

paradojas en big data: la ley de las poblaciones grandes, la paradoja de big data y las elecciones presidenciales estadounidenses de 2016", que más que hacer descarrilar al tren de big data, lo hizo girar 180 grados y lo mando al Hades profundo, "pa' los pagos de ande no se güelve", como cantaba José Larralde.

Los datos de cualquier encuesta tradicional parecen nada en relación con los que están disponibles gracias a big data. A la fecha, yo tengo 36.900 seguidores en Twitter, número que multiplica por algo más de 10 la cantidad de familias (3473) encuestadas por el Indec para medir una cuestión delicadísima como la pobreza. ¿David está ganándole a Goliat? Veremos. (Aprovecho la ocasión para un comentario: decirle "X" a Twitter me recuerda a cuando Prince pretendía que lo llamáramos con no sé qué signo del amor, que algunos incluso versionan como T, cosa que la mítica revista Humor interpretó literalmente, y pasó a llamarlo… "simbolito").

El contundente paper de Meng arranca con una pregunta clave, que remite a la confiabilidad de las predicciones y estimaciones: "¿Qué es más confiable: una encuesta científica hecha al 1% de la población y a la que responde el 60% de los encuestados, o una base de datos que cubre al 80% de la población?".

Una encuesta científica, de esas a las que Mayer-Schönberger y Cukier hacen bullying al condenarlas a una inminente obsolescencia, es una base de datos que tiene una estructura probabilística, vale decir, un nexo explícito entre los datos y la población que representa. Una base de datos de big data es… tan solo una base de datos. La enorme diferencia entre la muestra de una encuesta bien hecha y la de big data consiste en que, precisamente, la primera tiene una estructura, una forma sistemática de conocer cómo pasar de la muestra a la población. Cuando está bien hecha, una encuesta procede de una población y dice explícitamente cuál es el mecanismo que hace que alguien de la población aparezca en la muestra y cuáles son las probabilidades de que esto ocurra.

El muestreo simple al azar es algo así como la versión estadística del movimiento rectilíneo uniforme de la física del secundario: un ideal, un punto de partida. En este esquema, cada persona de la población tiene las mismas chances de aparecer en la muestra, y

que aparezca es independiente de que otra lo haga. Es como si hubiese una lista de personas (que conforman “la población”) y una moneda que se tira para cada persona: si sale cara, es encuestada y si sale ceca, no.

¿Cuál es la ventaja de proceder de esta forma, guiada por el azar? Que la muestra resultante es justa (no favorece a nadie en particular) y que aumentar su tamaño redunda en una ganancia en la precisión de lo que se pretende medir; una autopista suave y segura hacia la tierra prometida de los datos, esta vez sí. Salvo en condiciones extremas, tomar una muestra menor que toda la población implica aceptar una pérdida de precisión. Pero la matemática del muestreo al azar garantiza dos cosas. En primer lugar, que es posible calcular qué tamaño de la muestra hace que los resultados se mantengan dentro de cierto margen preestablecido de error. Este es un elemento clave de las encuestas, que, en el caso del Indec, da alrededor de 3743 para esta cantidad de encuestados. En segundo lugar, asegura que si se pudiese incrementar el tamaño de la muestra, la precisión aumentaría de modo acorde.

A pesar de sus ventajas, en la práctica el paradigma de muestreo simple al azar raras veces se usa en su forma pura: aunque se basa en un subconjunto pequeño de la población, es caro y difícil de implementar. A modo de ejemplo, supongamos que Carlos, que vive en el piso 18 de un edificio del populoso barrio porteño de Belgrano, salió sorteado para ser encuestado. Mirtha, la encuestadora, va a la torre en la que vive Carlos, toca el timbre, hace su tarea, y de ahí tiene que ir a la casa de Azul, que también salió sorteada y vive a doce cuadras. Mirtha nota que Carlos comparte departamento con sus dos hermanos y se tienta a encuestarlos a ellos también, y así evitar ir a lo de Azul. “Todos ganamos”, se dice, “encuesto a tres en vez de a dos y me ahorro el viaje”. De proceder de esta manera, tendrá más datos, pero… habrá roto el azar.

¿Y cuán grave es romper el azar? Gravísimo. De hecho, si Carlos es hijo de una familia acomodada que lo mantiene mientras estudia en Buenos Aires, hay altas chances de que sus hermanos tengan un estatus socioeconómico similar, si no idéntico. Dicho de otra forma, si reemplaza a Azul por los hermanos de Carlos, Mirtha recoge más

datos, pero no necesariamente obtiene más información. En el extremo, si Mirtha encuestase tres veces a Carlos, tendría tres observaciones, pero la misma cantidad de información que con un dato solo. El azar garantiza representatividad e independencia, pero a un costo operativo alto: a Mirtha no le queda otra que olvidarse de los hermanos de Carlos y patear doce cuadras para entrevistar a Azul. El muestreo al azar es operativamente caro.

Otro inconveniente asociado con el azar radica en que no asegura que la representatividad de una muestra para una población sea válida también para alguna subpoblación de interés. Por ejemplo, en una muestra de 100 personas al azar, no es raro que aparezcan muchos más varones que mujeres. Una idea tentadora es encuestar a la misma cantidad de hombres y mujeres, pero esto… rompe el azar.

Por este motivo, en la práctica las encuestas no son al azar. Las encuestas científicas modernas provienen de un delicado proceso que negocia las ventajas de pegarse al azar con las consecuencias de alejarse. Desviarse del azar es la kriptonita de las encuestas, ni hablar cuando no existe una forma de medir (menos aún, de corregir) los costos de apartarse. Y esto es lo que le pasa al tren de la alegría de big data que, casi por construcción, no sigue las reglas del azar y, si vamos al caso, ninguna regla.

El gran problema es que los datos de big data surgen de ver lo que hace la gente, de rastrear las “huellas digitales” que dejan las acciones de las personas en sensores, likes en Twitter, clics en Mercado Libre o imágenes en cámaras de seguridad. Por construcción –como decimos los estadísticos–, big data no tiene una estructura, no hay una forma obvia de atar una muestra a una población de referencia. En relación con el caso de mis seguidores en Twitter, no es claro a qué población refiere cualquier encuesta que yo haga más allá de… la población de gente que me sigue en Twitter vaya a saber uno por qué. A modo de ejemplo, en los días candentes de las elecciones presidenciales de 2023 había un fenómeno muy simpático. Todas las encuestas que alguien hacía en Twitter favorecían toscamente sus propias preferencias. Así, los kirchneristas mostraban que sus encuestas daban a Sergio Massa como claro ganador, y lo mismo, pero con signo contrario, ocurría

con los simpatizantes de Javier Milei.

En definitiva, los datos de big data no son más de lo mismo, no son versiones extra large de las encuestas profesionales. Si lo fuesen, los estadísticos deberían ponerse contentos como Bon Scott, o quizás tristes sabiendo que, como decían Mayer-Schönberger y Cukier, el automóvil de big data llegó para reemplazar al burrito cordobés de la estadística.

¿Tiene sentido hacer encuestas en épocas de IA?

Pero, como sospecharán, todo es mucho más complicado. Lo que Xiao-Li Meng hizo es "ponerles nombre y apellido" a los problemas que trae big data cuando, en su anarquía, se aleja del ideal del muestreo al azar.

Vayamos ya mismo a un resultado escalofriante que obtiene Meng en su trabajo. Al juntar todas las encuestas disponibles realizadas en los Estados Unidos para las elecciones de 2016, en las que resultó electo Donald Trump, había unas 2.315.570 personas consultadas. Es decir, agregar todas las encuestas que flotaban por los medios era el equivalente a haber hecho una megaencuesta entre un poco más de 2.000.000 de personas. ¿Mucho? Hmmm… Meng nota que, por esto de alejarse del azar, la confiabilidad de estos guarismos es comparable con la de una encuesta científica de… 400 encuestados. Dicho de otra forma, alejarse del azar tiene un costo similar a tirar el 99,98% de los datos a la basura.

Pongámoslo de otra manera. Supongamos que Hugo es responsable de esta megaencuesta de 2.315.570 observaciones y que, distraído por bailar desaforadamente con AC/DC de fondo, borra todos los datos. Desesperado, llama a Alberto, su entrañable amigo experto en estadística, y le dice "por favor, sacame de este berenjenal; vendo toda la colección de vinilos de AC/DC para que hagas una encuesta". De acuerdo con los resultados de Meng, si con esta plata Alberto hace una encuesta a 400 personas elegidas al azar, Hugo puede salvar las papas y garantizar la misma confiabilidad que con los datos que borró mientras sonaba (sí, adivinaron) "Highway to Hell".

Entender de dónde sale el resultado de Meng es complejo, ya que involucra alguna matemática más o menos sofisticada; pero haremos un esfuerzo para acercarnos a la idea principal. La pregunta es de qué depende la diferencia entre los resultados de una encuesta y los que se obtendrían de disponer de toda la

población. Sobre la base de una identidad matemática, Meng encuentra que la diferencia entre esas dos magnitudes depende solamente de tres factores, a los que llama

grado de dificultad del problema por medir,

cantidad de datos,

calidad de los datos.

Si bien uno puede hacerse una idea intuitiva de qué hay detrás de estos conceptos (que parecen bastante razonables), la fórmula de Meng es muy precisa al respecto.

El grado de dificultad es la dispersión del resultado por medir. Vayamos a un caso simple, dado por el ballottage en la Argentina, en 2023. Para simplificar el ejemplo, supongamos que solo se puede votar a Milei o a Massa, es decir, no hay voto en blanco ni impugnado. Obviamente, si permitimos el caso más realista, casi nada cambia; solo se complica el vocabulario usado. En el caso de Milei o Massa, en la situación extrema en la que toda la población vota a un candidato y nadie al otro, una encuesta de tan solo una persona permite adivinar el voto de toda la población. En esta circunstancia (extrema y absurda), el grado de dificultad es nulo, es la tabla del 1 de las encuestas. La cosa se complica si, como en la realidad, los votos están repartidos. Y acá viene lo interesante. No es difícil demostrar matemáticamente algo que ustedes quizás puedan intuir: que el máximo nivel de dificultad para una encuesta está dado cuando la proporción de votos es 50%/50%.

Este es un resultado viejísimo de la estadística –Meng no lo descubrió, ni mucho menos– y dice algo que a nadie le gusta escuchar: que la estadística es menos confiable cuando uno más la

necesita, es decir, cuando la elección está reñida.

La elección de Barack Obama en 2008 fue vista como un auténtico triunfo de las encuestas. El analista y escritor Nate Silver se coronó como el héroe de los datos al predecir acertadamente los resultados. Unos años después, cuando irrumpió Trump en la política, Nate Silver, los gurúes, las encuestas y enseguida las estadísticas pasaron a la categoría de villanos. Con el diario de mañana, la explicación es fácil y se relaciona con la fórmula de Meng. Obama venció a McCain por una diferencia de casi un 7,2%, mientras que la diferencia entre Trump y Hillary Clinton fue del 2,1% (favorable a Clinton). Entonces, de acuerdo con la fórmula de Meng, la elección de Trump fue más difícil de pronosticar que la de Obama. En el caso de Massa-Milei, el libertario ganó con el 55,65% de los votos; desde este punto de vista, las encuestas la tuvieron difícil.

El segundo elemento en la fórmula de Meng para determinar la confiabilidad de las encuestas es la cantidad de datos. Desde cierta perspectiva, es bastante obvio que más es mejor. Pero hay una sutileza, porque lo verdaderamente importante no es la cantidad de datos sino la proporción de encuestados en relación con la población. Es intuitivo pensar que cuanto mayor es esta proporción, mayor es la confiabilidad. Suena razonable.

El tercer elemento de la fórmula de Meng es la madre de todos los líos: la calidad de los datos. Es obvio que más calidad es mejor, suena casi a una frase de sobre de azúcar. Pero la fórmula de Meng dice que la calidad relevante para la confiabilidad de las encuestas no es un concepto vago sino muy preciso: se trata de la correlación existente entre a quién vota una persona y las chances de que esa persona aparezca en la encuesta y la responda. Sí, esto es mucho más bravo de entender que una frase motivacional de José Narosky (infantes, por favor, busquen en Google).

Un ejemplo nos ayudará a comprender esta cuestión. Hay una correlación entre el voto y la chance de responder si los votantes de Milei participan menos en las encuestas que los de Massa. Misma cosa si las encuestas se hacen en zonas históricamente afines al peronismo. O si a los votantes de Milei o a los de Massa les da vergüenza revelar su voto. Entonces, cuanto más se relacione el voto a un candidato con las chances de que sus votantes aparezcan

en las encuestas, más correlación hay entre esos dos fenómenos.

De acuerdo con la fórmula de Meng, esta es la única noción de calidad que importa a efectos de la confiabilidad de las encuestas. Y esta correlación es lo que el azar evita, porque si en verdad la encuesta está hecha al azar, no favorece ni perjudica espuriamente a los votantes de Massa ni a los de Milei; en el ideal del muestreo al azar, esta correlación debería ser nula.

Resumiendo, según Meng, una encuesta electoral es tanto menos confiable cuanto

más repartido esté el voto en la población (mayor dificultad),

menor proporción de la población se encueste (menor cantidad),

más alta sea la correlación entre quien es encuestado y a quién vota (menor calidad).

El efecto mariposa de las encuestas mal hechas

La auténtica caja de Pandora se abre cuando Meng pone a la fórmula a trabajar con la realidad. En sus palabras, la fórmula aplicada a los comicios de un país desencadena una especie de "efecto mariposa" de desconfianza: una pequeñísima correlación entre voto y chances de ser encuestado es capaz de provocar una brutal caída en la confiabilidad de una encuesta grande, aun "de big data".

Acá la cuestión clave es apartarse del azar. Si se toman el trabajo de leer la letra pequeñísima de las encuestas electorales, verán que explicita la cantidad de encuestados. Por ejemplo, una encuesta estándar de las que se usaron para medir la intención de voto entre Massa y Milei se hace con unos 1000 encuestados, como los 1017 de la encuesta de Espop mencionada antes. Y en letra aún más chica, casi ínfima, se habla del margen de error que tiene toda encuesta, como ya analizamos. Así, debajo de cualquier encuesta encontrarán una frase del tipo "error estadístico, $\pm 3{,}5$, para un nivel de confianza del 95%", frase que casi nadie entiende, incluso muchos que estudiaron estadística. Este tipo de frases indica el margen de error esperable por haber consultado a una muestra y no a toda la población.

Un punto importantísimo es que, en el contexto del muestreo al azar (o alguna estrategia que se le asemeje), el margen de error depende estrictamente de la cantidad de encuestados y no de la población. Entender con detalle esta afirmación resulta engorroso, pero, si apelamos a una analogía del notable paper de Meng, es fácil intuirla. Supongamos que tenemos una olla enorme con salsa y que nos interesa saber si está salada. Cualquier persona sensata revuelve, mete una cucharita, prueba y procede en consecuencia. Usemos un punto de vista estadístico para observar mejor. El primer paso es revolver bien la salsa, el segundo es elegir el tamaño de la cuchara, el tercero es probar. El equivalente estadístico de "salsa

bien revuelta" es el muestreo al azar, lo que garantiza que ningún lugar de la olla tiene más sal que otro. Y ahora empieza la magia. Si la salsa está bien revuelta, la confiabilidad de la prueba no depende del tamaño de la olla sino del de la cuchara: una sopera es más confiable que una de café. En nuestra analogía, el tamaño de la olla es el de la población, y el tamaño de la cuchara, el de la muestra. Y acá viene lo sorprendente: es el "revolver bien" lo que ata la calidad del procedimiento al tamaño de la cuchara y no al de la olla.

Ahora bien, es un tanto obvio que cuando la salsa no está bien revuelta, la relación entre el tamaño de la cuchara y el de la olla empieza a importar. A modo de ejemplo, supongamos una olla enorme y un cocinero que, de puro vago, echa la sal en la parte de adelante y no revuelve. Extrapolar cuán salada está la salsa a partir de probar una cucharadita tomada de la parte de adelante de la olla conlleva un error que es tanto más grande cuanto mayor sea la olla. Apelar a una cuchara más grande inserta en el mismo lugar no solo no resuelve el problema, sino que lo agranda: nos da una sensación de falsa seguridad, porque nos hace confiar aún más en un resultado erróneo: una cuchara mayor "tiene más datos".

La contribución clave de Meng es formalizar el ejemplo anterior: cuando nos apartamos del azar, la confiabilidad de una encuesta no depende solo de la cantidad de encuestados (la cuchara), sino también del tamaño de población (la olla). Y, peor aún, la fórmula de Meng muestra que una pequeñísima caída en la calidad (entendida como la correlación entre voto y chances de aparecer en la muestra) es brutalmente amplificada por el tamaño de la población.

En su puntilloso análisis de las elecciones estadounidenses de 2016, Meng muestra que una mínima subrepresentación de los votantes de Trump es capaz de causar una caída dramática en la confiabilidad de las encuestas. Y peor aún: esta caída resulta más contundente en estados con mayor población. De hecho, Meng muestra que los únicos tres en los que las encuestas resultaron relativamente confiables fueron Hawái, Washington DC y Vermont, de los más pequeños en el ranking de habitantes.

Meng llama a este fenómeno la paradoja de big data: ante la presencia de sesgos, "cuantos más datos, nos autoengañamos con

más seguridad". Sí, como si al usar una cuchara más grande en la parte más salada creyéramos más enfáticamente que toda la salsa lo está (y termináramos convencidos de que ese es su sabor). Eso equivale a decir que la cantidad de datos amplifica los problemas.

Una conclusión práctica que Meng extrae es aparentemente paradójica, máxime desde la perspectiva triunfalista de Mayer-Schönberger y Cukier y su comentario sobre los caballos de las encuestas y los autos de big data: posiblemente sea preferible prestar atención a encuestas muy bien hechas aun cuando sean de tamaño ínfimo. Porque –y aquí parafraseemos a Schopenhauer–, salir arando con la Ferrari está bueno si es en la dirección correcta; caso contrario, no es una autopista al cielo sino al averno de los datos, el de la muestra grande de datos incorrectos.

En la circunstancia apropiada, el "burrito cordobés" de las encuestas bien hechas, al que "la siesta parece darle / una paz que huele a miel", le gana a la liebre no por rápido sino por astuto. Inteligencia natural vence a artificial. Mi perro cazó una mosca.

Big data persigue al covid-19

El hallazgo de Meng no es una curiosidad intelectual ni un problema específico del ámbito de las encuestas electorales. Un equipo liderado por Valerie Bradley, de la Universidad de Oxford, que incluye a nuestro infaltable Meng, exploró recientemente este problema de la paradoja de big data en un ámbito delicado: las encuestas de vacunación de covid-19. Esta historia es realmente una de David versus Goliat.

Por las razones conocidas, la pandemia disparó una alocada carrera para trackear su derrotero. La comunidad de big data-machine learning se dio por aludida y, con la mejor de las intenciones, se sumó a esta noble causa de monitorear la evolución y la intensidad del fenómeno.

Así, los "Goliat" de esta narración son el proyecto Facebook-Delphi y la Encuesta de Pulso de Hogares del Buró del Censo (HPS, por sus iniciales en inglés) de los Estados Unidos. El primero es una iniciativa conjunta de Facebook y la Universidad Carnegie Mellon. Se trata de una encuesta online, implementada en Facebook, que logró 250.000 respuestas por semana, lo que dio un total de 4,5 millones de datos entre enero y mayo de 2021. El HPS recopiló 600.000 respuestas en período similar. ¿Big data? Posiblemente. En todo caso, en términos comparativos, una barbaridad de datos. Hasta aquí Goliat.

Nuestro "David" es el Axios-Ipsos Coronavirus Tracker (que bautizaremos AICT), una encuesta mucho más modesta, de solo 1000 respuestas por onda, que en el período analizado logró solo unas 10.000 observaciones. Si fuera una película para postadolescentes, comenzaría con una escena en que los grandotes (Facebook-Delphi y HPS) le roban el almuerzo y le hacen bullying al AICT, que, a la larga, logró recabar solo un 0,2% de los datos de Facebook-Delphi y un 1,6% de los de HPS. Pobrecito.

Las tres encuestas dan resultados muy dispares entre sí. Un punto importante es que la brutal cantidad de datos en las dos primeras encuestas tiene un error de medición ínfimo, asociado a la enorme cantidad de observaciones involucradas. En términos de la analogía de la sección anterior, ambas se basan en cucharones enormes, en comparación con la diminuta cucharita de café del AICT.

La "verdad verdadera", como diría mi hijo, quedó arriba de la mesa: allí la posó el Centro para el Control y Prevención de Enfermedades (CDC) estadounidense, que llevó un registro exacto del proceso de vacunación. Claro, lo hizo con el diario de mañana, pero es la "verdad de referencia" (ground truth es la terminología sajona usada en la jerga predictiva), un aspecto clave en el entrenamiento y reentrenamiento de modelos. Las encuestas intentan adelantarse al CDC, cuya información es lo que permite ver quién gana o pierde, como los resultados electorales, que revelan cómo les fue a los encuestadores. La conclusión es la que ustedes están sospechando. Delphi-Facebook y el HPS sobreestiman persistentemente las tasas de vacunación, mientras que las predicciones del modesto AICT tienden a ser correctas.

Y aquí entra en escena la bendita fórmula de Meng, que permite explorar el porqué de las fallas de los Goliat de las encuestas. De acuerdo con lo analizado antes, hay que poner el ojo en la calidad de los datos.

El estudio de Bradley encuentra desviaciones groseras con respecto al azar en las encuestas de Facebook-Delphi y el HPS. Ambas sobrerrepresentan a los adultos educados de piel blanca, a los demócratas y a la población urbana; subpoblaciones con tasas más altas de vacunación. Por el contrario, un cuidadoso tratamiento de ponderaciones mitiga estos sesgos en el caso del AICT. Una vez más, el ideal de muestreo al azar implicaría que ningún grupo más o menos proclive a vacunarse está obviamente sub- o sobrerrepresentado. Una cuenta grosera usando la fórmula de Meng da que las 250.000 observaciones de Facebook-Delphi contienen información equivalente a no más de 25 observaciones de un estudio hecho al azar. David sonríe.

Twitter e IA al rescate de las encuestas electorales

Los peligros que encuentra Meng refieren a un uso naïf de los datos, a tomar la información “de big data” como si proviniese de una encuesta. A fin del escrutinio político, no se trata de tirar los datos de las redes sociales, sensores o imágenes a la basura y seguir confiando tercamente en las encuestas tradicionales, sino de hacer un uso inteligente de ellos.

Las encuestas tradicionales tienen muchísimos problemas, y no es claro que el paso del tiempo y los avances de la tecnología no contribuyan a acentuarlos. Por ejemplo, las tecnologías de IVR (sigla inglesa de “respuesta de voz interactiva”) siguen ocupando un lugar importante en la recolección de datos. En criollo, IVR es esa tecnología que nos llama al teléfono de línea y con un mensaje pregrabado nos pide que respondamos las opciones ingresando números en un teclado. En el mejor de los casos, la rápida desaparición de los teléfonos de línea solo genera “no respuestas” que redundan en menos información y, en el peor, sesgos de cobertura, de los que espantan a Meng, ya que quienes insisten en retener un teléfono de línea no son una muestra al azar de la población. Por ejemplo, mi madre, octogenaria, si bien maneja muy bien el celular, insiste en conservar su teléfono de línea “porque me es más fácil pedir turnos médicos” (sic) y lo usa solo a tal fin. Entonces, es posible que la población remanente con teléfonos de línea sobrerrepresente a los mayores y deje de lado a los jóvenes, que, según mucha evidencia empírica, recientemente han tenido un rol crucial en los eventos electorales de varios países.

La pregunta obvia es si un uso inteligente de los datos de big data no permitirá complementar y eventualmente reemplazar a las encuestas tradicionales. Lo obvio, como hemos mencionado, es que los datos de las redes sociales son copiosos y baratos, casi por construcción (como también dicen los expertos en argumentación). Un estudio reciente del consultor argentino Luciano Cohan (y varios

coautores) explora el potencial de la IA aplicada a Twitter para el pronóstico de resultados electorales. Como bien cabía esperar, la cantidad de datos resulta espeluznante: el análisis de estos autores se basa en 110 millones de tuits, de 6,3 millones de usuarios. Cifra sideral comparada con los aproximadamente 1000 datos de una encuesta tradicional.

Un uso inteligente de big data requiere entender que la masividad de datos viene acompañada por un sinfín de cuestiones por resolver, para que los sesgos que preocupan a Meng no echen por la borda el esfuerzo. Lo óptimo sería que la fineza y precisión de las encuestas tuviera su correlato en Twitter. Nada más alejado de esto. Como dijimos, cualquier encuesta en esa red social está atravesada por el problema de que los usuarios no interactúan al azar, sino de acuerdo con patrones que pueden sesgar los resultados. Es decir, seguir a tal o cual, darle un like, retuitear o comentar no es algo que ocurra de manera fortuita, sino según patrones e interacciones sistemáticas. Pero, por otro lado, la naturaleza anárquica de las redes sociales invita a expresar libremente las opiniones. Entonces, es fácil adivinar si uno de nuestros contactos tiene preferencias por tal o cual candidato (si revisamos sus muros). Esta es la tarea que llevaron adelante Cohan y sus coautores: entrenar algoritmos para percibir, en tiempo real, las preferencias políticas y así predecir resultados electorales. Los resultados sugieren que, con el debido cuidado, las redes sociales incluyen información relevante que puede superar la performance predictiva de algunas encuestas tradicionales.

Lejos del análisis ingenuo, los problemas que trae aparejados el estudio de datos de redes son numerosos. En primer lugar, como señalan los autores, las redes están plagadas de bots que meten una considerable dosis de ruido. Otro clásico problema de este método es la necesidad de estandarizar el lenguaje, o sea, de realizar cierto esfuerzo computacional para estudiar léxico, sintaxis y semántica en los tuits a fin de predecir si un usuario prefiere a tal o cual candidato sobre la base de lo que escribe.

Un punto relevante que marcan estos autores es que posiblemente lo que importe no sea la opinión específica de alguien, sino la acumulación de opiniones. Este es un tema relevante y una

limitación obvia de las encuestas tradicionales. La información provista por cualquier encuesta estándar refiere a un punto en el tiempo. Por varias razones, esta información puede ser volátil o estar fuertemente influida por lo que respondan otras personas. Varios estudios indican que estos problemas de "ostracismo social" hacen que muchas personas no quieran exponer sus preferencias, fenómeno documentado, por ejemplo, cuando Trump fue elegido presidente en 2016: muchos votantes blancos y con educación formal tardaron en revelar su intención de apoyar al empresario por temor a ser víctimas de "represalias sociales" de parte de sus vecinos o conocidos, y se volcaron masivamente a expresarla cuando se dieron cuenta de que muchos de sus pares sociales lo hacían. Cohan y sus coautores mencionan que este ostracismo social y la posterior estampida en favor de Trump fueron un factor clave del fracaso de las encuestas tradicionales mientras el empresario mediático irrumpía en el campo político. Muchos analistas locales sugieren que algo similar ocurrió con el caso del candidato libertario Javier Milei. A la fecha de escribir este libro, el estudio sistemático de este fenómeno para las elecciones de 2023 es una deuda pendiente.

En esta cuestión, las redes sociales tienen una ventaja clave en relación con las encuestas sociales. Por su conformación, estas últimas son "de corte transversal", como se dice en la práctica. Es decir, para un momento en particular, se encuesta a unas 1000 personas, al azar o, como ya vimos con detalle, usando algún otro mecanismo que se le parezca. Para otro período, se encuesta a otras 1000 personas, y así sucesivamente. Lo ideal sería encuestar siempre a las mismas 1000 personas, de modo que las comparaciones fueran "de peras con peras", lo cual también permite estudiar la dinámica de las preferencias, si ciertos fenómenos tienen un impacto en que una persona esté dispuesta a cambiar o mantener sus opiniones. Es casi imposible estudiar estos fenómenos con datos de corte transversal, en que los respondientes cambian con cada encuesta.

Las encuestas de datos en paneles, en que a una misma persona se la encuesta en varios períodos, son costosísimas, en grado tal que su implementación práctica es imposible. El principal costo es que, si ya resulta difícil que una persona responda una encuesta, mucho

más lo es que se comprometa a hacerlo durante varios períodos.

Resulta que Twitter (y, en general, las redes sociales) no tienen este problema: más allá de bots y fakes (fotos trucadas, videos truchos, etc.), la gente tiende a sostener durante varios períodos sus cuentas y a vociferar a troche y moche sus gustos, opiniones e inquietudes. Cohan y sus coautores explotan esta crucial característica de las redes sociales para estudiar no solo lo que la gente opina en un momento, sino también la dinámica de estas opiniones: si se acumulan, si cambian, si son erráticas. Más específicamente, la acumulación dinámica de opiniones parece ser un factor clave en el uso de las encuestas "de big data" y una ventaja respecto de las encuestas tradicionales.

¿Veremos una mejora sustancial en la capacidad predictiva de las encuestas electorales provenientes de big data y la IA? Posiblemente sí, siempre y cuando las complicaciones se refieran a los problemas de captación de las encuestas tradicionales, que, más allá de lo de Meng, se puedan resolver con ciencia de datos. Pero en todos los casos es válida una advertencia, relacionada con el tema que desmenuzaremos más adelante, en el capítulo 9: en política, en economía y en las cuestiones sociales, las dificultades predictivas tienen que ver con los problemas de captar las preferencias de las personas (en lo que big data seguro traerá considerables mejoras) y también con que las cuestiones sociales tienen un componente muy difícil de conjeturar, por su naturaleza estratégica. Atacaremos este complejo tema, sí, páginas adelante. Por ahora, les dejo una pregunta, para que vayan practicando: ¿cómo puede ser que en épocas de big data, de IA, de algoritmos, de redes sociales, nadie sepa cuánto va a costar el dólar la semana que viene y, peor aún, que con el paso de los años no haya habido mejoras en la predicción este guarismo? Algo similar afecta a la política, al deporte y a varios hechos sociales.

Volver del futuro

Hay algo fundamentalmente contradictorio en repetir, como loro, que “las encuestas electorales no funcionan”, y que sigan ocupando un lugar prominente en la discusión pública. Bueno, no olvidemos que una de las máximas de las ciencias sociales es que la estupidez persistente es en sí misma un fenómeno social relevante. Si las tan cuestionadas encuestas se usan, es porque algo dicen; que no digan todo no significa que no digan nada. He aquí un resumen de ideas que tratamos en este capítulo:

La relevancia de las encuestas electorales depende decisivamente de la “honestidad” de los mecanismos que recaban los datos utilizados.

El mecanismo honesto por excelencia es el del muestreo al azar, que garantiza que la encuesta no favorece ningún resultado que no sea la verdad.

El muestreo al azar es caro y uno siempre tiene la tentación de abandonarlo; así, les abre la puerta a los enormes problemas que tienen las encuestas mal hechas.

Como herramienta, big data no necesariamente resuelve los problemas de las encuestas tradicionales (por lo menos, no por sí misma).

Pequeñas “deshonestidades” (sesgos) se amplifican cuando se dispone de más datos de la fuente errada; por eso, una encuesta de pocos datos pero bien confeccionada puede ser mucho más útil que una “de big data”.

Más datos no necesariamente implica más información. Y, peor aún: como muestra la paradoja de Meng, a veces la masividad solo conduce a errores más pronunciados, como le pasa a quien sale corriendo en la dirección equivocada.

Si en un futuro big data y la IA llegaran a solucionar las limitaciones de las encuestas, no será por la masividad de datos ni por la potencia de los algoritmos, sino porque facilitarán un uso creativo de métodos que permitan corregir sesgos.

3. Predicciones para la economía y la guerra

Si con la política la teníamos difícil, con la economía ni les cuento. De hecho, una vieja chanza dice que la economía se inventó para que la meteorología parezca precisa.

Una buena parte de la actividad predictiva se basa en buscar patrones y, como es de esperar, mucho de la práctica de los pronósticos se basa en complejos mecanismos que hurgan en datos. El problema con las cuestiones sociales es que los datos hablan de la realidad tanto como meten barullo. Peor aún, muchas veces nos dicen lo que tenemos ganas de escuchar.

Parte del rotundo éxito de la IA se basa en el poder que tienen las computadoras en encontrar recurrencias y relaciones esquivas a la percepción humana. Las historias de este capítulo muestran que uno debería ejercer cierta cautela o, mejor aún, apelar a alguna inteligencia “natural” antes de tirarse a la pileta de los patrones estadísticos. Siendo optimistas, argumentaremos que los mejores procesos predictivos tienen tanto de encontrar patrones como de razonar creativamente. Es decir, consisten en una adecuada combinación de inteligencia artificial y natural.

La maldición del Premio Martín Fierro Revelación

El Premio Martín Fierro Revelación destaca a jóvenes actores y actrices de la Argentina que sobresalen en sus comienzos. En el ámbito de las predicciones, este tipo de reconocimientos puede ser un predictor (y también un espaldarazo) de una carrera promisoria. ¿O todo lo contrario?

Una parte de la prensa (y también del público) instaló la idea de que más que un premio es una maldición: resulta una carga difícil de llevar y, cual tumba profanada de Tutankamón, exuda un conjuro que infesta a aquellos que lo ganaron, de quienes solo se puede pronosticar lo peor. Como evidencia, se citan los casos de Noemí Frenkel, Jorge Schubert, Daniela Fernández y otros ganadores del premio en su momento, aparentemente alcanzados por el maleficio; "nunca más se supo de ellos, / del caso no se habló más", como de los oficiales que vieron morir a Manuel Santillán, en la canción de Los Fabulosos Cadillacs. En la misma línea, en Estados Unidos se habla de "la maldición de la tapa de Sports Illustrated": quien aparece en la portada de esta popular revista de deportes atrae, más que un espaldarazo, solo malos presagios.

A los efectos del contrapunto, como en cualquier payada, podríamos pensar en un "Premio Viejo Vizcacha Revelación", que "premia" a lo peor de los medios. Y de existir la maldición del Martín Fierro Revelación, ¿no debería haber algo así como la "bendición del Viejo Vizcacha Revelación"? ¿No sería cierto que conductores papeloneros, locutores aparatosos y actrices soporíferas mejorarían sustancialmente sus performances luego de ser deshonrosos merecedores de este premio? ¿No sería el Viejo Vizcacha Revelación un predictor de éxito y un servicio a la comunidad?

Estamos frente a uno de los problemas más famosos y menos entendidos de la estadística predictiva: la falacia de la regresión a la media. Si la regresión a la media tuviese una sede física, e hiciera como esas panaderías que se jactan de su longevidad en el barrio,

en su entrada diría: “Confundiendo a científicos e incautos desde 1886”.

En el ámbito de la estadística, la idea de regresión surge de un estudio de Francis Galton publicado, sí, en 1886, en el que documenta copiosamente un fenómeno raro: a padres altos les suceden hijos más petisos y, al revés, padres muy petisos tienen hijos un poco más altos; la altura de las personas parece “tener una regresión al promedio”. Galton es uno de los padres de la estadística moderna y, en plan de explorar esta cuestión de padres e hijos, puso la piedra fundamental para muchas técnicas e ideas que todavía se siguen usando profusamente en ciencia de datos: regresión, correlación, la relevancia de la distribución gaussiana (o normal, en términos estadísticos: da esa forma de campana tan apacible en los gráficos), entre muchas otras.

Entendida al pie de la letra, la idea de regresión estudiada por Galton sugiere que, si un padre es muy alto, entonces debería predecirse que su hijo será más bajo, y lo contrario con padres petisos. Desde esta perspectiva, la maldición del Martín Fierro Revelación es una especie de profecía autocumplida: de actores o actrices tempranamente reconocidos por sobresalir con respecto al resto, solo se debería esperar que, mansa y obedientemente, retornen a la media. Y lo mismo, pero con signo contrario, debería vaticinarse de los “merecedores” del Premio Viejo Vizcacha Revelación: el tiempo los sacará del extremo inferior donde fueron encontrados y los devolverá al sitio que nunca debían abandonar: la media.

Entonces, ¿todo lo que sube tiende a bajar? ¿De los comportamientos extremos solo se puede predecir un sumiso retorno al centro? Sí y no. La confusión que muchos tenemos con la idea de regresión a la media se relaciona con no comprender en qué circunstancia ocurre o, peor aún, como casi todo en la vida, con el hecho de que las cosas retornen a la media es una cuestión de grado.

A fin de entender esta cuestión, comencemos pensando en un caso extremo. Supongamos que un grupo de alumnos en su primer examen universitario es evaluado con consignas del tipo “verdadero o falso”; para cada ítem los alumnos deben elegir y señalar una sola

de las alternativas, sin proveer una justificación. Supongamos que los alumnos responden tirando una moneda: si es cara, responden verdadero, y falso en caso contrario. Como todo es al azar, la nota esperada (en una escala de 0 a 10) para cada alumno es 5, es decir, que acierten la mitad de las preguntas. En esta circunstancia, los mejores alumnos (los que sacaron nota más alta) son sencillamente los que tuvieron más suerte. Si aislásemos a los mejores (y les diésemos el Premio Martín Fierro Revelación de esta universidad), veríamos que, si el examen se repite, la performance de estos alumnos debería bajar. ¿Por qué? Porque el azar, por definición, no se repite, si no, no sería azar. En resumen, estos alumnos han "retornado a la media". Y lo mismo habría ocurrido con los peores (¡merecedores del Viejo Vizcacha Revelación!), pero en la dirección opuesta.

Ahora vayamos al extremo contrario. Supongamos una prestigiosa y competitiva universidad, donde los alumnos estudian a más no poder y deben realizar un examen complejo, en el que tienen que pensar y acudir a todo el esfuerzo y talento que los ha acompañado en su derrotero y no, como en el caso anterior, a tirar una moneda. Acá habrá mejores y peores, pero el azar no interviene. Así, Mariela Sassano es reconocida con el Premio Revelación de esta universidad por su destacada performance en el examen. ¿Qué se espera que pase con ella en un segundo examen? Que repita su desempeño, como lo viene haciendo desde la primaria. ¿Retorno a la media? De ninguna manera, Mariela fue abanderada de la primaria, de la secundaria, hizo el mejor examen de ingreso y, por lo que se ve, sigue encaramada en la cima de los rankings.

Y aquí viene la trampa. En el primer caso, el único factor que determina las posiciones relativas de los alumnos es la suerte, no hay talento. El retorno a la media es simplemente un reflejo de que el azar no persiste. En el segundo caso todo es talento, la suerte no interviene, no hay retorno a la media porque el talento persiste y es lo único que explica estar arriba, abajo o en el medio.

La del militar que les grita a los soldados

La confusión con el retorno a la media viene porque en la vida real hay una adecuada mezcla de talento y suerte. Muy posiblemente, detrás de quien saca una nota alta en un examen haya una combinación de ambos. También sabemos que aun los más talentosos tienen un mal día o son víctimas de un episodio que los afecta negativamente, como la muerte de un familiar o de una mascota querida. Es decir, cuando suerte y talento aparecen mezclados, arriba hay un "combo" de talentosos y suertudos, y debajo, lo contrario.

En abstracto, el fenómeno de regresión a la media tiene que ver con qué sucede con un fenómeno dada una situación inicial, cuando está gobernado por un componente transitorio (que llamamos "suerte") y otro persistente (que llamamos "talento"). En este contexto, calculo que ya se habrán dado cuenta de cuál es la conclusión inevitable: el fenómeno de regresión a la media ocurre más intensamente cuanto más importante sea la parte transitoria en relación con la persistente. En el primer caso que contamos, el retorno a la media ocurre porque es todo azar; en el segundo no existe, porque hay pura persistencia. La vida real, como casi todo, se mueve en un interesante y peligroso intermedio.

El Premio Martín Fierro Revelación, salir en la tapa de Sports Illustrated y haberse sacado una nota altísima en el primer examen de la facultad son reconocimientos al resultado, a la mezcla de talento y suerte. Un único resultado no permite separar con claridad cuánto hubo de talento y cuánto de suerte, de ahí que en los Martín Fierro el fenómeno de la maldición del Revelación esté más sujeto a la cuestión del retorno a la media que el rubro Trayectoria, que reconoce a quienes se destacaron durante toda su carrera, de modo que la sociedad tuvo muchas oportunidades de separar adecuadamente factores fortuitos de persistentes, lo que explica que nadie hable de tal cosa como "la maldición del Martín Fierro a la

Trayectoria".

Pero la realidad complica aún más las cosas, ya que lo persistente no lo es tanto y la suerte no es tan loca. Así, algún alumno que en su primer examen obtiene una elevadísima calificación quizás aprenda qué funcionó y qué no en su estrategia de estudio y se vuelva una persona "más talentosa", dejando menos espacio para el azar. Dicho de otra forma, el fenómeno de regresión a la media no es estático, sino dinámico.

Tal vez la cuestión más preocupante de las confusiones en torno al fenómeno del retorno a la media es que no se trata de una cuestión causal, sino más bien de un artificio estadístico. A modo de ejemplo, aun cuando haya retorno a la media en el caso del Martín Fierro Revelación (por los factores que explicamos antes), no es que la carrera actoral de, digamos, Jorge Schubert se haya desvanecido por haber ganado este premio, sino por una mezcla de factores compatibles con haber retornado a la media, incluido que quizás él mismo haya elegido correrse del centro de su profesión (de hecho, ese parece ser el caso de Schubert, quien desde hace varios años se dedica a la escritura). En el caso extremo del examen que se resuelve tirando una moneda, de haber un premio a los mejores alumnos, que luego estos retornen a la media en exámenes posteriores no tiene nada que ver con ningún premio, sino con la modalidad azarosa a la que recurren para resolver los exámenes. Es decir, el premio no causa nada, simplemente acompaña un fenómeno de la realidad, como los ladridos del perro de la cuadra que corre detrás de todos los autos que pasan y se cree que los espanta.

En la práctica es muy difícil separar factores persistentes de temporarios, y es fácil caer en la tentación de atribuirle causalidad al fenómeno de retorno a la media. En su jugosísimo Pensar rápido, pensar despacio, el laureado psicólogo y economista Daniel Kahneman, recientemente fallecido, cuenta una historia que ilustra está confusión. Un duro militar le comentó que la técnica que notó más útil para entrenar a su tropa es gritar a los soldados cada vez que se equivocan. Palabras más, palabras menos, dice el militar: "Cuando hacían las cosas mal, les gritaba y después mejoraban; por el contrario, cuando los felicité si hacían las cosas bien,

empeoraron. Entonces dejé de felicitarlos y ahora solo les grito". Veamos esta jocosa anécdota desde la perspectiva del fenómeno de regresión a la media. El militar cree que los "malos" mejoran porque él les grita. Si la performance de sus soldados está dominada por factores temporarios (como el azar), pensar que los gritos mejoran ese desempeño es como creer que los peores actores mejoran por haberles dado el Viejo Vizcacha Revelación. Si el azar es un factor relevante, los "buenos" empeoran luego de haber hecho algo bien y los "malos" mejoran luego de una performance subestándar, independientemente de que el instructor grite, alabe, solo grite, solo alabe o haga cualquier cosa. Desde la óptica del instructor, habría que abolir el Martín Fierro Revelación y reemplazarlo por el Viejo Vizcacha.

La presencia de sesgos confirmatorios complica aún más el análisis de la regresión a la media. Es decir, la tendencia innata (honesta o no) de sentirse inclinado a confiar en datos que confirman lo que uno cree. Así, los defensores de la existencia de la "maldición del Martín Fierro Revelación" mencionarán los casos más remanidos (y, en verdad, no muy abundantes) de carreras truncadas, sin decir que el mismo premio fue recibido por Federico Luppi, Susana Giménez, Oscar Martínez y tantos más, que han tenido una carrera persistentemente exitosa, lejana de cualquier conjuro conspirativo.

Entonces, en relación con la temática de este libro, ¿es razonable predecir fracaso luego de una performance descollante? Dependerá de la mezcla de cuestiones persistentes y transitorias. Que algunos hayan dudado del talento de Messi y que mucho seudoemprendedor exitoso se autoatribuya las razones de sus logros solo habla de que las viejas falacias nunca mueren, como dijo una vez Milton Friedman.

Regresión a la media y el futuro del capitalismo

El evento que presentaremos muestra que, además de llevar a asignarle efectos causales al Martín Fierro Revelación, la confusión con la regresión a la media condujo a uno de los papelones predictivos más recordados de la historia de la economía.

Una parte importante de la tarea predictiva es contar narraciones convincentes. Y en disciplinas que tienen dificultades para apelar a experimentos (por razones tanto éticas como operativas), como las sociales, buscar todas las historias compatibles con los mismos datos es casi un instinto de supervivencia. Así, por ejemplo, el economista o politólogo de fuste siente la obligación de contemplar la posibilidad de que una piña en el ojo haya sido un cabezazo al puño, no por maniático sino porque, ante la imposibilidad de experimentar, construir historias coherentes es a veces la única forma que tiene de acercarse a las explicaciones y a los fenómenos causales; una cosa es experimentar con un fertilizante en distintas parcelas, y otra, con el capitalismo o el comunismo en distintos países.

En 1933, intentando indagar en el futuro del capitalismo, Horace Secrist terminó (honesta, inocente y un tanto involuntariamente) protagonizando un papelón predictivo de resonancias épicas que, en pleno auge de la IA y big data, viene al caso narrar a modo de recordatorio de los peligros que conlleva abusar de las historias y los datos.

La obra de Secrist, titulada El triunfo de la mediocridad en los negocios –les ahorro la mención en inglés–, es un volumen de 468 páginas, con 140 tablas y 103 gráficos, meticulosamente documentado, y con una alarmante conclusión: “La mediocridad tiende a prevalecer en la conducta de los negocios competitivos… Ese es el precio que trae consigo la libertad”. Es decir, al revés del caos que pronosticaba el marxismo con relación al capitalismo, los resultados de Secrist predicen que, libradas a la dinámica

capitalista, empresas con desempeños inicialmente dispares tienden a converger todas en un insulso promedio: las mejores empeoran, las peores mejoran y, a la larga, prevalece una especie de “empresa promedio” y mediocre.

Secrist extrajo sus conclusiones de la observación de 49 tiendas multirrubro estadounidenses, desde 1920 a 1930. Su puntillosa investigación nota que las de mejor performance en 1920 tienden a empeorar con el tiempo, y lo contrario ocurre con las peores. El mismo patrón aparece en otros negocios, como carnicerías, ferreterías y bancos. Secrist culpa a las fuerzas competitivas del mercado de esta indeseable tendencia a que todas las empresas se igualen y se comporten como el promedio.

La primera reacción fue una lluvia de elogios a un trabajo tan preciso y estadísticamente refrendado. Sin embargo, al poco tiempo, Harold Hotelling, notable matemático, estadístico y economista, señaló una falla garrafal en el razonamiento que sostenía las conclusiones de Secrist.

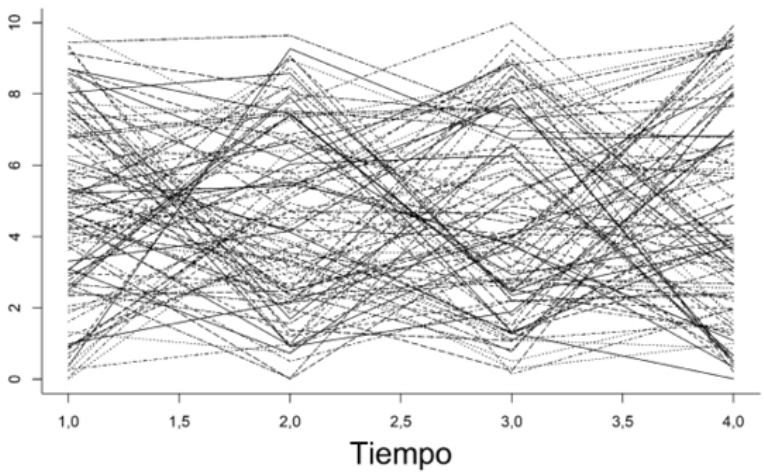
El artificio que detectó Hotelling en el estudio de Secrist es, precisamente, el de la espantosa cabeza de la regresión a la media asomándose. La aleatoriedad innata en la performance de las empresas hace que el retorno a la media sea nada más que una consecuencia trivial de haber aislado a las que al principio del análisis les fue bien, como los merecedores del Martín Fierro Revelación: fue el azar, y no las fuerzas del capitalismo, lo que empeoró las cosas para las que eran mejores al inicio.

Hotelling señala que si bien es cierto (de manera trivial, como discutimos) que el grupo de las inicialmente mejores empeora (y al revés, en el caso de las peores), de aquí no se sigue que todas las empresas convergen a comportarse como el promedio. Esto es difícil de entender, así que apelemos a un ejemplo.

Si siguiésemos a cualquier empresa en el período estudiado por Secrist, la veríamos deambular, hacia arriba o hacia abajo, al compás de los movimientos del azar que afecta la vida de cualquier persona o país. Y lo mismo ocurriría con cualquier empresa, a la que veríamos bailotear al ritmo de lo fortuito, yendo para arriba y para abajo. En todo período habrá empresas ganadoras y

perdedoras. En la siguiente figura, la 3.1, les muestro una simulación por computadora de la performance de 100 empresas artificiales (creadas con un algoritmo). Sigan con la vista cualquiera de ellas (cada línea representa a una empresa) y verán que aparece la fluctuación descripta en el párrafo anterior. ¿Hay convergencia al promedio? No, de ninguna manera. La brecha en la que oscilan las empresas nunca se cierra ni se abre, empiezan y terminan igual de heterogéneas. Lo único que hace el azar es revolearlas dentro de esta banda, haciéndolas ganar o perder y rotar en estos roles, pero asemejarlas, no.

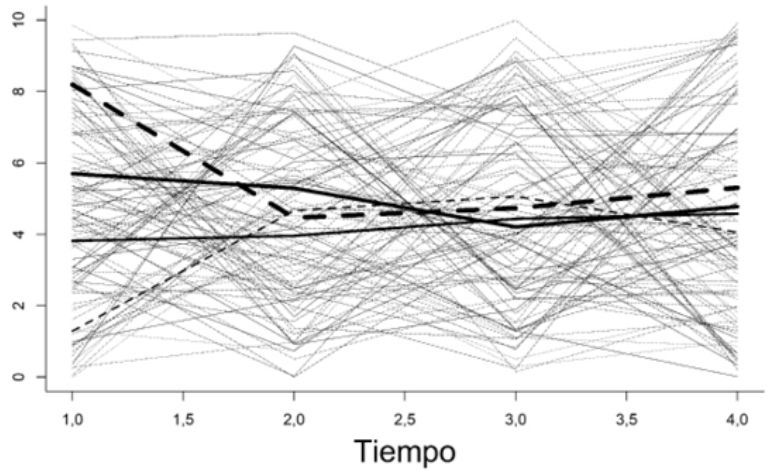
Figura 3.1



Ahora concentrémonos en cómo les fue a las empresas en el primer año y organicemos cuatro grupos: las 25 mejores (¡ganadoras del Martín Fierro Revelación de la Industria!), luego las 25 que le siguen, las otras 25 y por último las 25 peores (merecedoras del Viejo Vizcacha). Calculemos la performance promedio de cada grupo y hagamos un seguimiento de estos promedios en el tiempo. Es decir, en el segundo período calculemos la actuación de las mejores empresas en el período inicial, y así sucesivamente y para cada uno de los cuatro grupos.

Sí, sí, sí (como diría o cantaría Roberto Galán, el histórico y vetusto conductor televisivo). El siguiente dibujo es el que vio Secrist y el que usó para justificar sus predicciones en relación con la inviabilidad del capitalismo.

Figura 3.2



En este caso, la regresión a la media es nada más y nada menos que una ilusión óptica, que surge de haber agrupado a las mejores sobre la base de su performance inicial. No hay tal cosa como una igualación de las empresas al promedio, como vimos en el gráfico previo cuando observamos a las empresas sin agrupar. De hecho, como señala dramáticamente Hotelling en su crítica, si hubiésemos agrupado a las empresas de acuerdo con cómo les fue en el último período, habríamos obtenido la conclusión contraria: que el capitalismo desiguala la actividad económica.

Este es el yerro que detecta Hotelling: agrupando a las empresas a piacere uno puede atacar o alabar al capitalismo, como en ese chiste que circula sobre los contadores públicos: alguien pregunta cuánto es 2 más 2, un matemático dice “4”; un estadístico, “4, más menos un error no mayor a 0,0000001”, y un contador, “¿cuánto

querés que dé?”.

En tiempos de IA y big data, es importante no sobreinterpretar patrones o tendencias, porque, como dijimos, las viejas falacias –entre ellas, la regresión a la media–, nunca mueren. La inteligencia natural, como la de Hotelling, debería estar más alerta que nunca.

La camioneta de Ernesto y los tanques alemanes

Las historias anteriores muestran que esto de "andar hurgando en los datos" para adelantarse al futuro es cosa seria, y no dejan de tener un tono pesimista en relación con los peligros de interpretar de modo irrespetuoso los patrones. Y como este libro no abandona su pretensión de ser un manual de autoayuda predictiva, la historia a continuación tiene un claro dejo realista y optimista: las mejores predicciones surgen de un uso inteligente, activo, del análisis de datos. Nos tocará ir a la guerra, a ver cómo datos más inteligencia es una combinación explosiva, con el respeto que merece la temática por abordar.

Ernesto Campomar fue un querido amigo que se fue más temprano de lo que es justo. Chanta adorable, protagonista de miles de anécdotas desopilantes y de profesión inescrutable. Inquirido al respecto, respondía con un lacónico "empresario", sobre todo ante el público femenino. Una vez, en una reunión de amigotes, alguien le preguntó: "Perdón, Ernesto, vos decís que sos empresario, pero ¿en qué empresa trabajaste?". "Ah… no, en ninguna, yo soy empresario en general". Todavía nos reímos de la anécdota de "empresario en general", mientras se nos cae una lágrima al recordar sus disparates.

Una vez se apareció por el barrio con una pequeña camioneta, ploteada con el nombre de uno de sus "emprendimientos", y que, en letras grandes, decía "Móvil nº 7". "Para impresionar a la clientela", dijo, entre nuestras risotadas. Y ahora caigo en la cuenta de que, con relación a los episodios de Ernesto, todas las palabras usadas deberían estar entre comillas, porque Ernesto vivía entre comillas. La pregunta predictiva es la siguiente: a los ojos de sus clientes, ¿cuántas camionetas tiene Ernesto? De no conocer su "linaje", la respuesta obvia es "por lo menos siete". Esta anécdota nos deja ante las puertas de un sutil problema predictivo, pero en un contexto más dramático, que muestra que analizar datos es mucho más que

un simple ejercicio burocrático-computacional; en realidad, requiere una sana dosis de astucia.

Vayamos con la máquina del tiempo a la Segunda Guerra Mundial. El relato que sigue es una simplificación numérica de un suceso real. Los Aliados logran capturar cuatro tanques alemanes, y sus servicios de inteligencia se dan cuenta de que cada tanque tiene un número de serie, como la camionetita de Ernesto. A fin de simplificar el problema, supongamos que los tanques capturados tienen estos números de serie: 2, 7, 14 y 4. La pregunta clave coincide con la que nos hicimos en el barrio con respecto al emprendimiento de Ernesto: ¿cuántos tanques tienen los alemanes?

Obviamente, en plena contienda los Aliados no tenían la menor idea. O sí. Convengamos que los alemanes tenían como mínimo cuatro (los capturados). Y, tal vez, como Ernesto en tren de engatusar a sus clientes, tenían solo cuatro y a uno le pintaron "14", para confundir al enemigo. Ahora bien, si los tanques fueron enumerados del 1 en adelante y sin saltear números, la respuesta pasa a ser "como mínimo, 14". A lo que Dwight D. Eisenhower, comandante supremo de las fuerzas aliadas en Europa, posiblemente haya respondido: "No sea estúpido, me está dando la respuesta correcta a una pregunta que yo no le hice; yo no le pregunté cuántos tanques hay como mínimo, sino cuántos hay realmente desplegados por el Reich".

Sí, es un problema difícil. Pero los científicos oyen "difícil" y, máxime en épocas de guerra si también resuenan las palabras "urgente" y "relevante", en vez de huir, afilan el lápiz. Y el cerebro. Lo obvio es que, en la suposición de que los tanques fueron numerados sucesivamente sin saltear ninguno, la mejor estimación numérica es 14; más allá de 14, vaya a saber uno qué pasa; puede haber habido 14, 15 o, si vamos al caso, 500. O 1550, como conjeturaban los servicios de inteligencia de los Aliados, sobre la base de sus investigaciones secretas con métodos y fuentes más secretos aún.

Y aquí es conveniente detenerse para pensar en dos cosas. Primero, proponer como solución que la cantidad de tanques coincide con el máximo número de serie de los capturados (14, en este caso) es casi como creer haber resuelto un rompecabezas y que sobren piezas.

Fíjense que para llegar a la conclusión de que 14 es el número de serie más alto de los capturados, el único uso que se les da a los números de los tanques es para ordenarlos y ver cuál es el mayor. Y noten también que la cantidad de tanques capturados (4, en este caso) no juega ningún rol en la solución propuesta. O sea, en caso de proponer “14 tanques”, hay mucha información subutilizada:

los números de serie de los tanques capturados,

la cantidad de tanques capturados.

Sospechoso. En segundo lugar, esta estrategia siempre yerra por defecto, salvo que haya exactamente 14 tanques en total. Proponer el máximo de los números de serie de los tanques capturados es como insistir siempre en disparar a la izquierda del blanco; hay algo contradictorio en que uno pifie y lo haga sistemáticamente. En el capítulo 5 diremos que, en este caso, la predicción es sesgada hacia abajo. Uno podría argumentar: “OK, ya sé que hay un sesgo, pero no tengo ninguna información para corregirlo, solo sé que la cantidad de tanques es por lo menos 14; más allá de eso, todo es pura conjetura”. Bueno, acá es donde el ojo de la ciencia ve más allá de lo obvio.

Resulta que en los datos que “tiramos” o “barajamos” cuando computamos 14 (el resto de los números de serie y la cantidad de tanques capturados) hay información fundamental para saber cuán lejos está el máximo (14) del verdadero total de tanques alemanes. Misterio; es difícil, se los advierto. Ahora les tengo que pedir, ¡otra vez!, que se hagan un té y piensen un poquito. O que se relajen y vean cómo termina esta de Sherlock Holmes.

Bueno, ahí vamos. Recordemos los números de serie de los tanques capturados, ordenados de menor a mayor:

2, 4, 7, 14.

Si esos tanques fueron capturados al azar, es como si hubiesen estado encerrados en una bolsa, fueron bien revueltos, alguien sacó cuatro y salieron estos con los consabidos números de serie.

Pensando en abstracto, la mínima cantidad de tanques que podrían tener los alemanes es 0 (chocolate por la noticia), la mínima cantidad de tanques que sabemos que tienen es 4 (los que vemos) y la máxima es justo lo que queremos estimar. Ahora fíjense bien. Entre 0 y el tanque 2 hay un tanque que no vemos (el tanque 1); entre el 2 y el 4 no vemos el tanque 3; entre el 4 y el 7, dos tanques (el 5 y el 6) y entre el 7 y el 14 hay seis tanques que no se ven (8, 9, 10, 11, 12 y 13). Aquí debajo les muestro los números de serie de los tanques capturados, agregando el 0 y, al final, x, el total de tanques de los alemanes, que es lo que queremos saber. Debajo puse la cantidad de tanques faltantes entre los que se observan:

Números d	0	2	4	7	14	x
dad de tanques n	1	1	2	6	¿?	

Resulta que ahora hay dos cosas que no sabemos: la cantidad total de tanques, x, y cuántos tanques no vemos entre el 14 y x, magnitud que denoté con “¿?”. Y seguro se dieron cuenta de qué es lo que vamos a hacer: si podemos calcular “¿?”, se lo sumamos a 14 (más 1), y eso nos da x.

El promedio de la cantidad de tanques no visibles entre los capturados es

$$\frac{1 + 1 + 2 + 6}{4},$$

lo que equivale a 2,5; ese es nuestro candidato para “¿?”. A esta magnitud la llamaremos “distancia promedio”, es decir, en promedio, cuántos tanques no visibles hay entre los capturados. Entonces, si hubiera más tanques que 14, parece razonable que, en caso de haber un número de serie más alto que el 14, esperaríamos que estuviera a la derecha de 14, a 2,25 + 1 de distancia, como la distancia que separa, en promedio, a todos los tanques capturados, empezando desde la mínima cantidad (0) y yendo de forma ascendente al total, x. La predicción sensata del total de tanques es el máximo observado, más la distancia promedio entre todos los capturados, más 1,

$$14 + 2{,}5 + 1,$$

o sea, 17,5 tanques. Esta es la lógica detrás del famoso problema de los tanques alemanes, un clásico de la estadística predictiva. Ahí va la fórmula:

nº de tanques predichos = máximo número de serie + distancia promedio + 1.

Si bien derivar la fórmula es complicado (en el apéndice les dejo QR para acceder a alguna referencia, por si se mueren de la curiosidad; eso sí, afinen el lápiz… el que avisa, no traiciona), hay alguna intuición fácil de por qué la fórmula debería andar razonablemente bien. Pensemos un ratito. Supongamos que se capturaron 4 tanques, pero que todavía no sabemos qué número de serie tienen. Empecemos suponiendo que los alemanes solo tienen 4 tanques; entonces, necesariamente, los tanques capturados tendrían los números 1, 2, 3 y 4. Ahora supongamos que la cantidad de tanques capturados sigue siendo 4, pero resulta que el total de tanques que tienen los alemanes es 20. Ergo, los números de serie de los tanques capturados deberían estar en el intervalo entre 1 y 20, y si fueron capturados al azar, entre ellos debería haber bastante más espacio que cuando la cantidad total era 4. No nos privemos de dar otro salto y llevar el caso más al extremo. Si Alemania tuviese 1000 tanques y capturamos 4, los números de serie (que ahora estarían entre 1 y 1000) aparecerían con valores mucho más espaciados. ¿Ya se dieron cuenta del truco?

Se los cuento: para un número fijo de tanques capturados, cuanto mayor es el total de tanques alemanes, más espaciados deberían aparecer los números de los tanques capturados. Y acá viene la táctica de “ingeniería reversa”: si cuando la cantidad de tanques es más alta, los números de serie de los capturados están más espaciados, cuanto más espaciados estén los números… ¡más alto tiene que haber sido el total de tanques! Entonces, se espera que la cantidad de tanques sea el máximo número observado más la distancia promedio que separa a todos los tanques capturados. Lo

que acabamos de hacer es estimar la distancia promedio entre los tanques (2,5) y agregársela al máximo valor observado (más 1), como si fuese el penúltimo. Maravilla de las maravillas.

Y si ni la matemática ni la intuición los convencen, podríamos hacer una simulación con la computadora, como haremos repetidas veces en este libro (el capítulo 5 se ocupa del uso de simulaciones para predecir). Pero como estoy relatando un evento pasado, otra vez tenemos a mano el "diario de mañana" de la guerra; podríamos leerlo, levantar el velo histórico y ver cómo le fue realmente a la fórmula propuesta usando datos reales.

Ya dijimos que los servicios de inteligencia aventuraban una cifra de 1550 tanques alemanes. El método analizado en esta sección (y aplicado a los tanques capturados en la real realidad, no los 4 que usamos para simplificar el caso) predijo que los alemanes tenían 327 tanques. Cuando los Aliados entraron en Alemania y encontraron los puntillosos archivos del temible Albert Speer (criminal de guerra y contra la humanidad, además de ministro de Armamento del régimen nazi), se dieron cuenta de que los alemanes tenían 342 tanques. Un poquito más de lo que había predicho la fórmula, una barbaridad menos de lo que sostenía la "inteligencia".

Hay moraleja: ojo con lo que dicen los datos a simple vista. A los gritos, los de nuestro ejemplito dicen "14 tanques". Un incómodo susurro dice "debe de haber más", y la voz precisa de la ciencia propone una forma de aprovechar información aparentemente inútil (la distancia entre los tanques observados) para corregir el sesgo. Este es un claro ejemplo de uso inteligente de datos y lógica demoledora. La tarea predictiva seria tiene tanto de buscar patrones como de pensar con precisión y creatividad.

Puestos a estimar cuántos móviles tenía el emprendimiento de Ernesto, la respuesta más razonable habría sido "ninguno", porque seguro que ni la camionetita con el número 7 era de él. En algunos casos, la inteligencia barrial domina a la artificial.

Volver del futuro

Si algo tienen en común estas historias predictivas en el marco de lo social es que los mejores pronósticos requieren un tipo de inteligencia sutil, que no es obvio que sea sinónimo de artificial. Predecir el comportamiento social con solo extra- o interpolar lo que se observa es como pretender inferir las leyes de la física y la electrónica a partir de nuestro bingeo de series en Netflix. Estas son algunas ideas clave de este capítulo:

El fenómeno de regresión a la media, que llevó a Secrist a pronosticar la irrelevancia del capitalismo, es un ejemplo de que analizar datos no es un mero ejercicio mecánico-computacional, sino que requiere de cierta dosis de inteligencia, porque un uso irresponsable de los algoritmos predictivos es proclive a dejarse llevar por ilusiones ópticas estadísticas.

En todo caso, el yerro del ejemplo de Secrist consiste en reportar solo los patrones que sostienen una conclusión (el fin del capitalismo) e ignorar otros que surgen de los mismos datos y que sugieren exactamente lo contrario.

Elegir un patrón de datos para justificar una conclusión establecida de antemano es como lanzar un dardo a una pared y luego dibujar círculos concéntricos alrededor del punto exacto donde quedó clavado el dardo, para argumentar que uno dio en el blanco, como en un dibujo que una vez me mostró Javier Joaquín, gran ilustrador argentino

La historia de los tanques alemanes es optimista y muestra el verdadero desafío de la actividad de predecir con datos y algoritmos. El análisis naïf de los datos de las numeraciones de los tanques alemanes no permite ir más allá de predecir que su cantidad es el más alto de los números de serie observados.

Solo un análisis inteligente muestra que para predecir una magnitud de interés (la cantidad de tanques) es relevante predecir primero otra cantidad (la separación promedio entre los números observados) y luego actualizar la predicción inicial.

4. Predecir con probabilidades: pronósticos y pronosticadores

Habrá sido en octubre de 2022, en alguna de las charlas que suelo dictar para alumnos de colegios secundarios sobre el universo de los datos, los algoritmos, las certezas y las dudas. Yo estaba hablando de los temas de este capítulo, y les pregunté: "¿Alguien sabe con certeza quién va a ganar el Mundial?". Por lo general me responden con un silencio respetuoso, pero esta vez un rubio flaquito me dijo: "Argentina, estoy seguro". Le retruqué: "Y si estás seguro, ¿por qué te parece que se va a jugar el Mundial, aunque vos sepas de antemano lo que va a suceder?". Y me aventuré más: "Si es tan obvio, ¿cómo es que no sos capaz de convencer al mundo entero de este resultado?".

Sí, el pibe acertó, seguro que apelando a una mezcla honesta de pasión y deseo. Pero ¿predijo correctamente? Con el diario de mañana, sí: Argentina ganó. ¿Y con el de ayer? Que el Mundial se haya jugado y que hayamos estado con el corazón en la boca hasta el último segundo de la final muestra que una predicción correcta es mucho más que acertar un resultado. Los pronósticos que funcionan son los que cambian cursos de acción antes de que las cosas pasen. Son los que nos inducen a cargar con un paraguas incluso si no se ve una nube, a dejar de comer porquerías después de que el médico ve nuestro análisis de laboratorio, a vender, a comprar, a estudiar más para un parcial. Predecir es argumentar, es analizar datos, es explicar, es separar señal y ruido: predecir es lidiar con el azar y la complejidad.

En el capítulo 1 discutimos que dar con las chances de que algo ocurra es una tarea crucial para producir pronósticos. En este capítulo veremos que muchas veces calcular chances es un evento complejo y también que un manejo inteligente de las probabilidades permite mejorar significativamente la capacidad predictiva.

La maquinola de Persi Diaconis

Lamentablemente, aun para eventos mundanos, calcular probabilidades es en general una tarea compleja. Para ilustrar esta dificultad déjenme contarles algunos pormenores relacionados con calcular chances del, en apariencia, más inocente de los eventos fortuitos sobre el que venimos hablando: lanzar una moneda. Y en relación con esta cuestión, el matemático Persi Diaconis dijo: “Si hay tanto problema en analizar esta situación común, imagínese el lector las dificultades que encontraría en interpretar los supuestos sobre el azar en cualquier análisis econométrico”. Auch.

Persi Diaconis es todo un personaje. Si con individuos bastante menos interesantes se han hecho películas, no se explica por qué no hay series (de varias temporadas) sobre su vida azarosa (tal vez la única con genuino derecho a ese calificativo). Hoy Diaconis es un prestigioso matemático, profesor de la Universidad de Stanford. Pero el derrotero que lo llevó a este sitio empieza con un niño fascinado por la magia, que dejó la escuela secundaria para unirse a una troupe de magos itinerantes que actuaban en cafés y barcos. Su inmensa curiosidad y su poco ortodoxa formación lo impulsaron a empezar dubitativos estudios universitarios de matemática en el turno noche.

Aun luego de convertirse en un prestigioso intelectual de las probabilidades, Diaconis nunca abandonó su interés por los juegos. Por el contrario, muchos de sus estudios indagan sobre la naturaleza azarosa en este tipo de situaciones. Cuidado, lo de los juegos no es una muestra de liviandad o irrelevancia, todo lo contrario: Diaconis se vale de ellos para explorar lo más profundo del conocimiento humano y las matemáticas. Búsquenlo en Google, no se lo pierdan. Diaconis tiene aspecto tanto de matemático loco como de mago bufón que podría fascinar por horas a chicos y grandes. ¡El mismo nombre “Persi Diaconis” parece un apodo artístico, como Fu-Manchú, Dynamo o Tu Sam!

Uno de sus objetos de estudio fue, precisamente, el azar involucrado en el lanzamiento de monedas. La frase de Diaconis en el primer párrafo de esta sección aparece en las conclusiones de un tan fascinante como esotérico estudio que indaga en algo en que en apariencia no tiene sentido indagar: ¿es realmente cierto que los lanzamientos de monedas son justos para dirimir algunas situaciones, como quién comienza el juego en un partido de fútbol? En ese caso, ¿las chances de que salga cara son exactamente iguales a las chances de que salga ceca?

¿De dónde sale la aleatoriedad al lanzar una moneda? Si tomamos una moneda que está apoyada sobre una mesa con la "cara" hacia arriba, la levantamos con cuidado unos milímetros –de modo que quede paralela a la mesa– y la soltamos, la moneda caerá indefectiblemente con la cara hacia arriba; no parece haber nada aleatorio en esto. Por supuesto, muchos dirán: "Eso no es lanzar una moneda". Los lanzamientos razonables involucran colocar la moneda arriba de la uña del pulgar, que a su vez está encerrado por el dedo índice, y luego levantar el pulgar con un movimiento brusco que impulsa a la moneda para que ascienda y dé vueltas por el aire; tras lo cual, gravedad mediante, comienza su descenso para ser atrapada con las dos palmas, como quien aplaude, o, para involucrar más azar aún, da contra un piso duro, donde rebota algunas veces para luego recostarse en su posición final.

Diaconis dice en su trabajo que, si uno tiene control preciso del ángulo inicial en el que está la moneda al ser lanzada, y de en qué lugar y con qué fuerza se le da el golpe inicial para revolearla, "el lanzamiento de monedas es pura física, no hay aleatoriedad". Es decir, controlando estos parámetros uno puede predecir con exactitud si saldrá cara o ceca. A tal efecto, inquieto como nadie, Diaconis mandó construir una máquina lanzamonedas que le permitió controlar la posición inicial, dónde se le pega a la moneda para revolearla y con qué fuerza, y mostró que no es difícil configurar estos parámetros para que la moneda salga siempre como uno quiere. Algunos hacen pan con masa madre, otros tejen una bufanda; a Diaconis le resulta más simpático ocupar su tiempo libre de otro modo.

Pero en la práctica los lanzamientos ni son tan triviales como

levantar la moneda un poquito ni surgen de la maquinola exacta de Diaconis. Todos hemos visto, en los mundiales de fútbol, al árbitro que muestra ceremoniosamente una moneda a los capitanes de los equipos para que comprueben que no hay ninguna trampa (como una moneda con dos cecas), hace el gesto mecánico que describí más arriba, y luego observa cómo la moneda revolotea cual mariposa y cae cara o ceca para luego decir quién saca y quién elige arco. Obvio que "puede fallar" (lo decía el recién recordado ilusionista Tu Sam cuando sus trucos no funcionaban), como en un cotejo entre Colombia y Paraguay en la Copa América de 2016, en que la moneda cayó… ¡de canto!

Y acá viene la magia, de esa que le gusta a nuestro amigo Persi. Apelando a una batería de fórmulas físicas, experimentos y cámaras de video que pueden reproducir en cámara hiperlenta (de modo de facilitar las mediciones), Diaconis y su equipo muestran que, en las condiciones normales en las que una persona cualquiera revolea una moneda, el lanzamiento favorece un poquito la cara que estaba arriba al inicio. Más concretamente, el fascinante estudio muestra que, si al lanzar la moneda la cara que estaba hacia arriba era "ceca", las chances de que luego del lanzamiento salga ceca son del 0,51. En otros términos, el lanzamiento no es neutral. La explicación es finísima y se relaciona con las condiciones iniciales (como las que veremos con detalle en el capítulo 6) y las leyes de la mecánica clásica, como discurre minuciosamente Diaconis en su trabajo (dejo la referencia en el apéndice; con algo de física del secundario, un poco de matemática y confianza, se entiende).

Imaginemos qué podría hacer un árbitro inescrupuloso con este conocimiento. Supongamos que juegan Boca y River, que el árbitro es de Boca, que sabe que Boca prefiere sacar y que este resultado se corresponde con que la moneda salga ceca. De acuerdo con el trabajo de Diaconis, para favorecer a Boca, debería revolear la moneda poniéndola en el pulgar con la ceca mirando hacia arriba. Más o menos, de 1000 lanzamientos 510 saldrían ceca y 490, cara, favoreciendo así a la parcialidad xeneize. Diaconis es el primero en reconocer la naturaleza chacotona de la conclusión y dice explícitamente que "para las monedas lanzadas al aire, los resultados clásicos de independencia con probabilidad del 0,5 son bastante sólidos", para desilusión de más de un referí bombero que

pensaba hacerse unos pesos con este truco.

Pero el punto de Diaconis no es este resultado, sino el hecho de que un análisis profundo de un fenómeno aparentemente tan obvio deschava un sesgo que se da de patadas con lo que uno cree acerca del funcionamiento de las cosas.

La verdadera conclusión que extrae Diaconis, y la que importa a los fines de pensar el rol del azar en los pronósticos, es la siguiente. Si asignar chances al más obvio de los eventos azarosos es complejo, ¿qué queda para la economía, la meteorología, las finanzas o el deporte, que involucran acciones que van mucho más allá de las que se pueden representar con una maquinita con una palanca y un resorte? Muchos de los problemas de impredecibilidad que atañen a estas disciplinas tienen que ver, fundamentalmente, con la enorme complejidad que conlleva medir el azar cuando interviene una multitud de eventos que confluyen en que llueva, entre la pelota en un tiro libre o una sequía afecte a la liquidación de divisas de un país.

Después de discutir el sorprendente trabajo de Diaconis, flota la sensación de que estas pequeñeces son juguetes intelectuales de quien tiene demasiado tiempo libre. En el capítulo 2 nos dimos una auténtica ducha fría de realidad, cuando examinamos que pequeñísimos sesgos (como los de la moneda de Diaconis) en las encuestas electorales explican por qué, aun en tiempos de big data e IA, es muy difícil hacer pronósticos políticos.

Agitado, no revuelto: el problema de Monty Hall

Del ejemplo de Diaconis aprendimos que es difícil asignar probabilidades. Del que sigue, aprenderemos que a veces no solo es difícil sino contraintuitivo. El problema de Monty Hall, que indagaremos a continuación, arrasa como tsunami contra el sentido común de casi todas (no es un error de tipeo, la cuestión de género no es menor, como veremos) y es interesante discutirlo, porque muestra que muchas veces lo difícil se camufla muy bien con lo trivial. Muchos yerros predictivos tienen como origen estas confusiones.

He aquí el problema de marras (las ganas que tenía de escribir "marras"…). Todo sucede en Trato hecho, un exitoso programa de entretenimientos de la televisión estadounidense conducido por el tal Monty Hall. Hay tres puertas cerradas, una de ellas tiene un auto detrás, y las otras dos, una cabra cada una. El participante, que llamaremos Aldo, gana si acierta la puerta detrás de la cual está el auto (¡suponiendo que le interesa más ganarse un auto que una cabra!). Aldo no sabe dónde está el auto y tiene que elegir una puerta. Luego Monty abre una de las que no eligió Aldo, muestra que detrás de ella hay una cabra y deja la puerta abierta con el pobre bicho ya expuesto ante las cámaras. Y le pregunta a Aldo si quiere insistir con su elección inicial o cambiar de puerta. El juego termina mostrando lo que hay detrás de todas las puertas. Y acá viene la pregunta: ¿Qué le conviene a Aldo? ¿Cambiar de puerta o quedarse con su elección original?

Casi todo el mundo cree que Aldo no gana nada si cambia de puerta, porque las chances de ganarse el auto una vez que Monty abrió una puerta y le mostró que había una cabra siguen siendo de 1 en 3. De hecho, hubo una compulsa popular en relación con esta cuestión, y el 87% de los encuestados respondió que Aldo no ganaría por cambiar su elección original. Mal de muchos…

Develemos el misterio de una buena vez: A Aldo le conviene

cambiar su elección inicial.

Y ahora, más que explicarles este contraintuitivo resultado querría invitarlos a pensar cuáles pueden ser las rutas que uno puede elegir para convencerse o, al menos, quedarse tranquilo. La primera alternativa es creerme, porque le he dedicado toda mi vida a esto y se supone que sé, y por eso escribo libros de estadística, machine learning, pronósticos y probabilidades, ¡qué tanto! La segunda es tomar lápiz y papel y hacer la demostración. Es matemáticamente simple, pero intuitivamente rebuscada; en un ratito la vamos a explorar, no se desesperen.

Más allá de la demostración matemática, pueden sentarse a pensar, y si les parece que no es cierto que a Aldo le conviene cambiar, enviarme una nota diciéndome que estoy equivocado o que estoy loco. Pero me toca advertirles que, en este caso, estarían reaccionando como los miles de lectores de la revista Parade cuando, en 1990, la matemática Marilyn vos Savant expuso la solución al problema. Horda que incluyó a varios con doctorado en probabilidad y estadística, muchos tildándola desde loca hasta inocente.

Otra cosa que pueden hacer para convencerse de que a Aldo le conviene cambiar es valerse de la vieja LGN del capítulo 1; es decir, jugar ustedes y ver qué pasa. Si buscan online (en el apéndice dejo un enlace), hay disponible una versión del problema de Monty Hall. Los invito a que jueguen repetidas veces. Prueben no cambiar la elección inicial, prueben cambiar siempre, prueben cambiar al azar, y sorprenderse con el resultado: tienden a ganar si cambian de elección. Y si se sienten aventureros como Marilyn vos Savant, pueden escribir un código para simular este problema; es un hermoso ejercicio para aprender a programar (consejo: si, precisamente, quieren aprender a programar, no hay nada como buscar un lindo problema). Si saben algo de programación, anímense, les va a venir muy bien. Por las dudas, en el apéndice, el QR les permitirá encontrar unas líneas de código en R y también en Python para simular el juego.

Y si no los convencí con mi prueba por autoridad ni con la computadora, intentemos razonar un poquito. Es delicado entender por qué, ahí vamos: ajústense los cinturones que se viene la

turbulencia. Esta explicación está adaptada de un hermoso artículo de Steve Pinker, cuya referencia también queda en el apéndice.

Primero, una cosa. No importa dónde está el auto ni qué puerta elige Aldo, Monty siempre puede abrir una puerta donde hay una cabra y que no es la puerta que eligió Aldo. Por ejemplo, si Aldo eligió la puerta 1 y el auto estaba ahí, Monty abre cualquiera de las otras dos, donde hay una cabra. Pero si el auto está en 2, Monty abre la 3, y si el coche está en la 3, abre la 2: siempre hay una puerta que no abrió Aldo en la que hay una cabra.

Supongamos primero que Aldo elige la puerta 1. Ahora pensemos qué pasa si Aldo elige no cambiar su decisión: si el auto está en la 1, gana, y pierde si está en la 2 o en la 3. Si Aldo no cambia, las chances de ganar son de 1 en 3. Y acá un puntazo: estas chances no dependen de la puerta que abra Monty. Piénsenlo.

Sigamos suponiendo que Aldo elige la puerta 1, pero que ahora elige cambiar. Si el auto está detrás de la 1, el participante pierde. ¿Y si estaba en la puerta 2? Bueno, en esta circunstancia Monty habría abierto la puerta 3, en cuyo caso Aldo debía cambiar de la 1 a la 2… ¡y ganar! Y si el auto estaba en la 3, misma cosa: Monty habría abierto la 2, Aldo cambiaba y ganaba. Entonces, si cambia, las chances de ganar son de 2 en 3, el doble de las chances de no hacerlo, que eran de 1 en 3. Fíjense que, para cerrar el argumento, solo falta repetir lo que les comentaba recién: que este razonamiento es independiente de qué puerta elige Aldo. Entonces, a Aldo le conviene cambiar.

Como se habrán dado cuenta, la clave es que Monty no abre una puerta al azar sino “una que no abrió Aldo pero que no tiene un auto”. Ahí está la madre de los líos. Estoy seguro de que esta explicación no los va a convencer fácilmente, van a tener que pensarla muchas veces, sacar lápiz y papel y, tal vez, consultar varias fuentes.

El mensaje principal es que, muchas veces, cuando se trata de asignar probabilidades, la peor circunstancia aparece cuando este proceso se disfraza de fácil y oculta algo que se da contra nuestras intuiciones, como en el caso de Monty, que parece fácil pero no lo es.

La otra cuestión que esta historia pone sobre la mesa es qué es lo que nos convence. A algunos (espero que a pocos) habrá sido mi autoridad como experto (Aldo cambia porque yo te lo digo; sí, "sho", que me dedico a esto). A otros, habrá sido la explicación lógica, que vieron que simple no es. A otros, la experiencia, jugar el juego varias veces, probando quedarse o cambiar. Este es otro punto importantísimo para este libro: no hay una forma única de convencerse de que un pronóstico vaya a funcionar.

El efecto Mirtha Legrand y las probabilidades condicionales

Sí, la historia de Monty Hall es compleja, a más de uno le da dolor de cabeza. Pero hay situaciones más cotidianas que también conducen a yerros predictivos peligrosos.

A fin de ilustrar cuán difícil es interpretar los resultados de algunas manipulaciones probabilísticas en la vida real, les cuento esta historia. La señora Mirtha Legrand es una icónica y longeva conductora, la más importante de la historia de la televisión vernácula, al punto tal que la gente y los medios se refieren a ella como "la Señora". En algún momento, adivinar su edad era uno de los ejercicios predictivos por excelencia de las sobremesas de los argentinos.

En fin. Cuenta la señora que, en repetidas ocasiones, le sucede lo siguiente. Sube a un avión, se sienta, se ajusta el cinturón de seguridad y, al rato, se sienta al lado otro pasajero –a quien llamaremos Ovidio– que le dice: "Uy, señora, qué suerte que viajo a su lado". "¿Por qué?", retruca la diva, entre halagada y curiosa. "¡Porque ahora hay menos chances de que el avión se caiga!". "¿Por?". "Y… si es raro que un avión se caiga, y también raro que yo me encuentre con usted, que un avión se caiga y yo me encuentre con usted ¡es rarísimo!".

Veamos con detalle el razonamiento de Ovidio. Hay varias cosas por estudiar, o varios "eventos", como se denomina en probabilidad a las cosas que pueden ocurrir:

que un avión se caiga, viajen Ovidio, Mirtha Legrand, Pocho La Pantera o quien sea,

que se caiga un avión en el que viaja Ovidio,

que se caiga un avión en el que viaja Mirtha Legrand,

que se caiga un avión en el que Ovidio sube y se da cuenta de que está Mirtha Legrand.

Tranquilos, no vamos a sacar ninguna cuenta, solo vamos a pensar.

Ovidio razona de esta manera. Es muy raro que un avión se caiga. Es también muy raro que me encuentre con Mirtha Legrand. Y es muchísimo más raro que me encuentre con Mirtha Legrand en el avión y que, además, se caiga. ¡Zafamos!

Si es verdad que lo que le preocupa a Ovidio es que se caiga el avión donde él viaja, no debería importarle mucho si a su lado está Mirtha Legrand, Pocho La Pantera o un amigo del tío de una compañera del secundario que toca la quena en una banda tributo a Kiss. Entonces, si bien es cierto que es extraño que él se encuentre con la señora en un avión, la cuestión no es esa, sino si el hecho de que viaje la diva de los almuerzos altera la probabilidad de que se caiga el avión en el que viaja Ovidio.

Que Ovidio (o cualquiera de nosotros) viaje en un avión no parece alterar las chances de que ese avión se caiga. Ahora bien, ¿de que se caiga un avión en el que viaja la señora? Es discutible. Por un lado, alguien podría argumentar que las chances de que haya un accidente aéreo en una línea comercial dependen de cualquier cosa menos del nombre y apellido de quienes viajan como pasajeros, amén de que las chances de que un avión se caiga son bajísimas (1 en 11.000.000 por año, según el sitio de NOVA, el programa de ciencia de PBS). Más allá de lo minúsculo de esta cifra, tendríamos margen para creer que, a sabiendas de que viaja Mirtha Legrand, la empresa tomará medidas extras de seguridad para evitar la potencial mala prensa de que haya un desperfecto en un vuelo que

lleva a una celebridad a bordo.

Entonces, de ser verdadera la primera conjetura (la más razonable, obvio), si bien es muy extraño que Ovidio se encuentre con la Legrand y que el avión se caiga, el hecho de que viaje Mirtha no altera las chances de que haya un accidente fatal. Nada por ahí. Pero sí puede ser cierto que las chances bajen si ocurre la segunda conjetura, en que la presencia de la diva impulse a mejorar la seguridad, y ahí Ovidio hace bien en ponerse contento, pero no por las razones que aduce.

En términos técnicos se trata de discernir si la probabilidad de que un avión se caiga difiere o no de que se caiga dado que viaja Mirtha Legrand. La primera es una probabilidad y la segunda una probabilidad condicional. Confundir ambas probabilidades es la madre de la mayor parte de los fracasos predictivos. Más aún, explotar las probabilidades condicionales es la llave secreta de casi todos los aciertos. No es exagerado decir que los enormes éxitos predictivos de machine learning y la IA dependen de separar inteligentemente las probabilidades de las probabilidades condicionales. Todo pide un ejemplo, ahí vamos.

La épica librería Daniel

En plan de explorar la potencia de las probabilidades condicionales para la predicción (y los problemas de ignorarlas), déjenme compartirles un relato de mis días de estudiante. Es el año 1983 y yo soy un estudiante de Economía en la Facultad de Ciencias Económicas de la Universidad de Buenos Aires (UBA). Tras pasar doce años en una escuela barrial, la UBA me resultaba un enérgico y saludable caos, en el cual la supervivencia dependía de procesar correctamente el tsunami de información que corría entre los alumnos. En ese contexto, una decisión crucial era en qué cátedras anotarse para las distintas materias. Los buenos alumnos buscábamos aprender, y los malos, aprobar a cualquier costo.

Y en ese caótico proceso alguien me dijo: "Andá a lo de Daniel y fijate en las planillas". El tal Daniel era el CEO y el peón de una fotocopiadora de la calle Uriburu. Y era el arquetípico porteño chanta: entrador, de edad indescifrable, hablaba con la mayor desenvoltura sobre cualquier tema, en particular de todas las cátedras de la facultad, como si él fuese profesor o un alumno exitoso de todas. Su local brindaba un valioso servicio a la comunidad estudiantil: en el minúsculo y químicamente sospechoso local (por la tinta que usaba, descifrar los jeroglíficos egipcios era una nimiedad en comparación con leer sus fotocopias) tenía pegadas en las paredes unas enormes planillas con todas las cátedras de todas las materias de la facultad, con el nombre del profesor y un numerito que indicaba el porcentaje de alumnos aprobados en cada curso. Nunca supimos el origen de esos guarismos, o antes bien preferíamos no saberlo: preservar ese arcano da cierta cuota de tranquilidad, como nos pasa con las hamburguesas de la cancha.

Cual protorred social, a la información de la planilla se le agregaban anotaciones espontáneas de los alumnos (del tipo "inaprobable" o "está loco") y también los comentarios orales del

propio Daniel, quien, fiel a su estilo, opinaba con soltura del dictado de materias que iban de Biometría a Teoría y Técnica Impositiva IV, pasando por Lógica y Metodología de las Ciencias o Estructura Social Argentina.

Incluso si confiamos en los datos de la fotocopiadora ("tengo mis contactos", acotaba Daniel, guiñando un ojo), exploremos qué uso se les puede dar. Supongamos que lo que le interesa a un alumno son las chances de aprobar. Marcos y Vanina están pensando en hacer Contabilidad III y no saben en qué cátedra anotarse, y en lo de Daniel ven que el profesor Oroquieta aprobó al 70% de los alumnos.

Marcos capea la facultad como puede. Hace como que trabaja en el negocio familiar, juega al fútbol, a la playstation, al TEG; el estudio nunca fue lo suyo, la carrera no le interesa, no tiene en claro por qué la sigue estudiando y prevé que el semestre que viene tendrá la misma actitud. Vanina es hacendosa, le pone mucha garra, sabe que tiene una oportunidad de oro y no la desaprovecha: con mucho esfuerzo y algo de talento, jamás reprobó ninguna materia.

¿Cuáles son las chances de que Marcos o Vanina aprueben la materia dictada por el profesor Oroquieta? A falta de cualquier otra información, un 70%, siempre y cuando la de Daniel sea confiable, como, por ahora, estamos suponiendo.

Ahora, indaguemos un poco más. Supongamos también que Oroquieta es un buen profesor y, por añadidura, un evaluador justo (después veremos cómo cambiar el análisis si esto no fuera cierto). Vanina, responsable y calculadora, inteligentemente hace jugar a su favor la información que ella tiene y Daniel no: se sabe una buena alumna. Y si Oroquieta es justo, sabe que sus chances de aprobar superan el 70%. ¿Por qué? Precisamente porque tiene noción de que no es una alumna promedio y que, como siempre a lo largo de toda su carrera, ella hará el máximo esfuerzo. Marcos, poco proclive a lo analítico, se anota con Oroquieta porque cree que las chances de aprobar son 70%, como jura la planilla de Daniel. Termina el semestre, Vanina agrega otra materia a sus logros y Marcos, un nuevo aplazo.

He aquí un punto central de la actividad predictiva: lo que importa

para la predicción no son las probabilidades, sino las probabilidades condicionales. Dicho de otra forma: no importan las chances (no condicionales, como se dice en la jerga), sino las chances que dependen de cualquier información que permita afinar la puntería.

Vanina se da cuenta rápidamente de este truco. Sabe que un 70% es la probabilidad "no condicional" de aprobar, es decir, que refiere a "un alumno" de los que tomaron la materia de Oroquieta con anterioridad. Ahora, ella explota esta información y le agrega algo que sabe: ella no es una alumna cualquiera, es de las que trabajan y saben cómo manejar sus recursos. En términos técnicos, Vanina parte de la probabilidad no condicional (el 70%) y la actualiza hacia arriba, con información que ella tiene y Daniel no. Vanina sabe que el 70% se refiere a las chances de que un alumno apruebe. Pero lo que a ella realmente le importan son sus propias chances de aprobar, para lo cual las probabilidades condicionales son las que valen. Vanina no sabe cuáles son estas chances, pero sí que superan el 70%.

Marcos no aprovecha esta condicionalidad. Toma como idénticas las probabilidades de que alguien apruebe con las de que él lo haga. Si se toma en cuenta que no hará mucho esfuerzo, que se distraerá con cualquier cosa, su pasado de fracasos y su reticencia al cambio, las chances condicionales de que él apruebe no son del 70% sino un índice mucho menor. Y esto en grado tal que quizás sea cierto que repruebe, como efectivamente ocurrió.

Entonces, a los fines predictivos, lo que realmente importan son las probabilidades condicionales, las que explotan todo tipo de información que pueda utilizarse, como hizo Vanina, que tomó el 70% de Daniel como un punto de partida. Un grueso error de principiantes en la cuestión de las proyecciones es no condicionar en información relevante, como Marcos.

Y ahora es cuando tenemos que volver al profesor Oroquieta. Supusimos que era bueno y justo. ¿Y si fuese completamente arbitrario, tanto que sus exámenes y su posterior corrección funcionaran como surgidas de tirar un dado (¡de diez caras, como los de los juegos de rol!)? Bueno, en esta circunstancia las chances de que Vanina o Marcos aprueben son el 70% original, no hay actualización posible, ese 70% no depende del esfuerzo ni de nada

por el estilo. ¿Y si Oroquieta fuese un profesor misógino que piensa que las mujeres tienen que ir a lavar los platos y refleja estas ideas decimonónicas en sus correcciones, perjudicando a las alumnas? Las chances aumentan para Marcos y bajan para Vanina. En definitiva, aprecien ustedes que la forma en la que reacomodamos (o no) la probabilidad inicial, no condicional de aprobar es, nuevamente, tomando las chances de base y procesando la nueva información.

Mucho de la esencia del campo de machine learning tiene que ver con métodos que realizan esta tarea de forma eficaz, sobre la base de datos abundantes y algoritmos. Reseñar estos avances será la tarea del capítulo 7. A modo de ejemplo, tengamos presente que los sistemas de recomendaciones de plataformas como Netflix, Amazon o Mercado Libre se basan en esta idea de "retunear" probabilidades para ajustarlas a casos concretos. Por ejemplo, si al 80% de los suscriptores les gusta Seinfeld, Netflix debería explotar que, por mi edad, el barrio en el que vivo y mi historia televisiva, las chances de que a mí me guste son mucho mayores, y lo contrario sucedería con mi hijo, que opina que las andanzas de Jerry, George, Elaine y Kramer son ñoñerías de viejos como yo. Veremos que son los algoritmos los que se encargan de partir de probabilidades brutas (marginales, como se dice en la jerga) y actualizarlas a las condicionales para mejorar la predicción.

Y una idea más. Cuando de un evento no se aprende nada acerca del otro, se dice que son independientes. Por ejemplo, que mi apellido tenga cuatro letras tal vez no diga nada acerca de si tengo más o menos chances de que me guste Seinfeld. En este caso, que a alguien le guste Seinfeld (o no) es independiente de la cantidad de letras del apellido; condicionar en ese evento no conduce a ninguna mejora predictiva. Calculo que habrán deducido que, para dos eventos independientes, la probabilidad de que algo ocurra es idéntica a la probabilidad de que lo mismo suceda dado el otro evento.

El Robin Hood del barrio de Núñez: azar y talento

Lo conté en uno de mis libros anteriores pero, “como viene al caso” (así decía el enorme Juan Carlos Calabró cuando quería repetir una anécdota), déjenme recurrir a esta vieja historia para ilustrar un importantísimo punto con relación a esto de predecir bien o mal: en la realidad, el azar viene mezclado con lo sistemático, con lo predecible, con el talento.

Un día nublado en Mar del Plata llevé a mi hijo y mis sobrinos (por entonces de unos 7 años) a un sitio donde se practicaba arquería, con arco y flecha de verdad. Luego de unas breves instrucciones, los niños comenzaron la práctica, con resultados dispares.

Mientras tanto, yo me mantuve al margen, con una actitud entre displicente y altanera. El instructor, guiado por la cordialidad y el desafío, me preguntó si quería probar. Intenté negarme, pero no lo logré. Luego de un par de indicaciones, que juzgué inconducentes, tomé el arco, extendí hacia atrás la cuerda, cerré un ojo, apunté y… di exactamente en el blanco, para asombro de los niños. Dejé el arco, saqué mi celular del bolsillo, caminé hacia el blanco y tomé una fotografía que atestigua mis logros con la arquería. Jamás volví a tirar con arco y flecha, y nunca lo volveré a hacer. Tengo que proteger mi récord mundial de rotundo 100% de efectividad en arquería; una vez en la cima, todo camino es cuesta abajo, soy el Robin Hood del barrio de Núñez.

Y aquí aparece un punto central de la actividad predictiva: los eventos que son interesantes de predecir son el resultado de una mezcla de cuestiones predecibles, que obedecen a una causa o patrón, y otras que, a falta de mejor nombre, las llamamos “suerte” (de la buena, claro). En el caso del lanzamiento de la flecha, la parte sistemática tiene que ver con seguir las indicaciones del instructor, manejar correctamente el brazo, apuntar bien, etc. La parte fortuita tiene que ver con la falta de control preciso del cuerpo y las manos, lo errático de la visión, el viento y otras

condiciones azarosas que ocurren durante el disparo.

Mi único lanzamiento fue también una mezcla de suerte y pericia. Ahora bien, ¿será posible separar estos ingredientes? De la posibilidad de identificar cuánto de mi performance tuvo que ver con mi talento en la arquería y cuánto de suerte dependerá otra posibilidad, la de predecir qué ocurrirá con mis próximos disparos. No estiremos innecesariamente la conclusión de esta discusión inútil: lo mío fue un 100% suerte. El punto crucial es cómo podemos llegar a esa conclusión obvia, no la conclusión en sí.

A tal fin, es importante separar dos ideas. Una es más bien del mundo de los objetos (el rifle o el arco y la flecha), y otra, del mundo de las acciones (los disparos o lanzamientos). Volviendo al lenguaje del caso del rifle, el disparo es lo que resulta de manipular el rifle. Rifle, por un lado; disparos, por el otro. No me confundan gordura con hinchazón.

La teoría de las probabilidades tiene jerga específica para esta distinción. El rifle sería una variable aleatoria y los disparos son realizaciones de variables aleatorias. Ninguna definición de variable aleatoria será lo suficientemente precisa si no apelamos a funciones medibles, espacios muestrales y otros tecnicismos de la matemática profesional. Entonces, acudiremos a alguna idea simple. Una variable aleatoria es una variable que puede tomar resultados al azar. Por ejemplo, lanzar una moneda admite dos resultados razonables: cara o ceca. Entonces, la variable aleatoria es el lanzamiento de la moneda, que puede salir cara (con chances del 0,5) o ceca, con idénticas probabilidades. El lanzamiento específico de la moneda y el resultado que sale se denominan realización de una variable aleatoria. En el caso del arco y la flecha, la variable aleatoria es el hecho de lanzar con el arco y la flecha, que puede terminar en varios resultados: los puntos donde terminará la flecha si la disparo. La realización de esa variable aleatoria es un disparo en particular.

Y esta discusión nos deja la pelota picando para la definición de un actor fundamental de este libro: los datos. Sí, los datos son las realizaciones (o resultados) de las variables aleatorias, y estas, los mecanismos detrás de ellos. Predecir bien es acertarle a la máquina, analizar datos para ver qué hay de sistemático y qué de errático. Si

retomamos la analogía del rifle, los disparos son los datos y el rifle el mecanismo. La tarea de la ciencia de datos moderna es hacer el recorrido inverso: partir de los disparos (los datos) para aprender del rifle (el mecanismo) y predecir su comportamiento.

Pronósticos versus pronosticadores: “El argentino que predijo la crisis”

Este libro es una invitación a abandonar la visión binaria que solemos tener del futuro. Hay una diferencia entre que un evento ocurra o no (llueve, sale cara, aumenta el precio de una acción, o no) y sus probabilidades de ocurrencia. Lo primero es dicotómico, mientras que lo segundo no lo es, ya que puede abarcar un rango de valores, desde 0 hasta 100. Por lo tanto, si bien un evento puede ocurrir o no, hay una gran diferencia entre decir “no tengo la menor idea de si va a suceder” y especificar sus probabilidades de ocurrencia.

Y aquí viene una pregunta fundamental para este libro: si un meteorólogo dice que las chances de que llueva son de un 20% y otro, que son de un 90%, y al final del día no llovió, ¿quién es mejor predictor? Piénsenlo, piénsenlo.

Bueno, solo con esta información, es imposible saberlo, porque es una posibilidad contemplada en ambos pronósticos, si bien con distintas chances.

Si alguien afirma que las probabilidades de que una moneda salga cara son del 50%, ya hemos dicho que validar o refutar esa predicción es casi imposible basándose en el resultado de un solo lanzamiento. Una primera alternativa para evaluar al predictor es apelar, una vez más, a la LGN del capítulo 1: si nos dijeron que la chance de que salga cara es del 0,5, lancemos la moneda muchas veces (no una) y veamos si es verdad que en la mitad de los lanzamientos sale cara. Esa sería la vía empírica u objetiva de evaluar al predictor, que consiste en estudiar su historial predictivo.

La otra es la subjetiva, que –insisto– no es sinónimo de “caprichosa” o “arbitraria”, sino que equivale a decir que hay una explicación. Entonces, una forma alternativa de evaluar al predictor es pedirle que nos explique por qué cree que las chances de que salga cara son

del 0,5.

En la práctica ocurre una mezcla de procedimientos, y los pronósticos tienen lugar en ámbitos más dramáticos que lanzar una moneda. Así, cada tanto se oye hablar de “el argentino que predijo la crisis”, de que tal encuestadora acertó el resultado de las elecciones o del analista que pronosticó qué selección iba a ganar el Mundial.

La discusión que mantuvimos en este capítulo sugiere que la verdadera evaluación no es del pronóstico sino del pronosticador. Un buen pronosticador es alguien que revela sus métodos, sus fuentes, sus procedimientos (la parte “subjetiva”) o resiste la prueba del tiempo (la parte “objetiva”, relacionada con la LGN). Si 37 “analistas” se paran junto a una ruleta y cada uno dice un número distinto, del 0 al 36, uno de ellos le va a pegar, y a nadie se le ocurre venderlo como “el gurú de la ruleta”. Prueben soltar números al azar parados junto a una ruleta y verán que cada tanto van a acertar, no por sus dotes analíticas ni por sus conocimientos profundos de la probabilidad, sino por pura suerte.

Muchas veces, detrás de “el argentino que predijo la crisis”, hay un chanta que sistemáticamente dice que va a haber una crisis y, como un hecho estilizado de la macroeconomía es que toda economía fluctúa, cada tanto acierta, igual que en el ejemplo de la ruleta. La misma analogía es aplicable a “la encuestadora a la que mejor le fue en las elecciones”, porque es medio obvio que, si las predicciones difieren, una es la mejor de todas.

Los pronósticos que importan son los que nos llevan a alterar nuestro curso decisorio. Una cosa es que una persona cualquiera diga que un familiar tiene pocas chances de sobrevida, y otra que lo diga un médico que acaba de examinarlo con detalle. Es la experiencia y la sabiduría del médico lo que nos altera el curso de acción, no su predicción en sí. Una vez más, lo que importa no son los pronósticos sino los pronosticadores.

Y por eso este libro tilda de buen pronosticador a alguien que puede mostrar una historia de pronósticos acertados, o argumentos o explicaciones que sustenten esos pronósticos. Cualquier otro tendrá que seguir participando, ya sea pronosticando o explicando. Detrás

de quien “le pegó” a la crisis, a qué país ganó el Mundial o a qué número salía en la ruleta puede haber un suertudo con particulares habilidades comunicacionales, un farsante con un artilugio o un analista honesto.

“En Dios confiamos; el resto, que traiga datos”, dice una conocida chanza de las redes sociales, parafraseando la leyenda del billete de un dólar (In God we trust). En términos del ámbito de los pronósticos, la frase sugiere que un buen pronosticador debería venir con un buen historial predictivo o con explicaciones convincentes detrás de sus vaticinios. O también con alguna autoridad de la buena (ser Dios, en algún sentido relevante), pero de esa que surge de una apropiada mezcla de experiencia, sabiduría y comunicación.

Volver del futuro

La moraleja de este capítulo es que, si bien la presencia del azar molesta (y mucho), ciertos aspectos de algunos fenómenos son predecibles, como sus chances de ocurrencia. La pregunta correcta en el ámbito de las predicciones no es si aciertan, sino si las predicciones sirven para algo, aun sin pegarle por completo. Las historias de este capítulo giran en torno de estas ideas:

Predecir probabilidades es casi la esencia de la actividad predictiva.

Asignar chances puede ser una tarea compleja, como ilustran las historias relacionadas con Persi Diaconis y Monty Hall.

Como dice Diaconis, si la cuestión es compleja para eventos aparentemente inocentes, como lanzar una moneda, mucho más lo será para la economía, la política o la meteorología, en que el azar cumple un rol preponderante.

Un uso inocente de las probabilidades puede conducir, en el mejor de los casos, a predicciones inefectivas o, en el peor, a yerros sistemáticos. Confundir probabilidades con probabilidades condicionales es una de las fuentes más habituales de estos fracasos, como ilustra la anécdota del avión y Mirtha Legrand.

Un uso inteligente de las chances permite refinar la capacidad

predictiva, en particular, sobre la base de probabilidades condicionales, como en la historia que contamos de predecir aplazos en la facultad.

Tal vez la idea más importante de este capítulo es que una cosa es pegarle a un resultado y otra, predecir sistemáticamente. Lo primero requiere un poquito de suerte. Lo segundo, una audaz mezcla de análisis, datos e inteligencia.

5. ¿Errar es humano? Errores de pronóstico y el rol de las computadoras

Un tema que flota a lo largo de este libro es una cuestión que denominaremos la paradoja de la predicción: si algo se puede predecir con altísima precisión, es porque no era muy interesante haberlo predicho. Ante la omnipresencia del azar y la complejidad, la predicción sin error no existe; predecir implica lidiar con algún tipo de error. No hay mejor forma de evitar errores que no predecir nada, como cuando "se cuidan los zapatos / andando de rodillas", según cantaba Julio Sosa.

Entonces, el auténtico desafío de las predicciones no consiste en eliminar errores sino en negociarlos. Este capítulo se ocupa de dos cuestiones. Comenzaremos por analizar qué significa error de predicción, lo cual nos depositará en "el" problema predictivo por excelencia: la negociación entre sesgo e imprecisión, los dos sentidos en los en que un pronóstico puede fallar. Luego nos meteremos con otro aspecto fundamental de la predicción moderna: el rol de las computadoras en simular la realidad.

La balada de Laura y Susana: errores de pronóstico

En el secundario Laura era aplicada (tal vez el más escolar de los adjetivos). Así, llegaba a los asaltos (jóvenes de la sala, busquen en Google o pregúntenles a sus abuelos) vestida de punta en blanco, a la hora acordada y con una pasta frolla recién horneada. ¿Predecible? Como pocas. Susana era lo contrario: caótica, de letra ininteligible, podía pasar del 10 al 1 según le viniese en ganas estudiar o, sencillamente, se acordase de hacerlo. Infaltable en los bailes, aparecía a un horario incierto, determinado por sus vacilaciones en relación con el maquillaje y el outfit, y con un paquete de papas fritas comprado a las apuradas en un kiosco; por completo impredecible.

Este relato nos enfrenta a una confusión típica que atañe a la predictibilidad de varios fenómenos. Aparentemente, Laura es el ejemplo arquetípico de alguien predecible y Susana, lo contrario. Pero deberían empezar a sospechar que hay algo raro en predecir que alguien es impredecible, en línea con la paradoja de Epiménides, esa del filósofo cretense del siglo VI a. C. que dijo que "todos los cretenses son unos mentirosos". Si lo que dijo es cierto, el buen Epiménides no miente, de modo que no era cretense (una contradicción), y si es falso, entonces hay cretenses que mienten (Epiménides), otra contradicción. Ahora desátense el cerebro y sigan leyendo.

El problema es confundir lo impredecible con lo errático. Laura es estable, Susana es errática, condiciones que las acompañaron en su historia escolar y parecen continuar en un futuro cercano. Los fenómenos erráticos son difíciles de predecir en un sentido, pero sencillos en otro. De Laura, se sabe con precisión la hora a la que arribará al baile, cómo vendrá vestida y hasta qué calificaciones obtendrá en sus pruebas. De Susana, que aparecerá en algún momento, con algún snack comprado de apuro y que sus notas oscilan disparatadamente entre 0 y 10.

Laura (estable y regular) es predecible en dos sentidos, mientras que Susana (estable pero irregular) lo es en uno. Si nos focalizamos en el horario de llegada al baile, de Laura se puede predecir que llegará a la hora programada, con un margen de error de unos pocos minutos, que pueden deberse, por ejemplo, a que el taxi que la lleva encontró más o menos tráfico. Con Susana es otra cosa. Puede aparecer una hora antes con intenciones de instalarse en la habitación de la dueña de casa del baile y "tunearse" para la ocasión, o también una hora más tarde, si elige emperifollarse en su casa.

La predictibilidad del comportamiento de Susana y sus disparates es tan estable como la prolijidad de Laura. O sea, Susana es predeciblemente impredecible, para mal de sus padres y de Epiménides. En términos técnicos, y con relación al horario en que Laura y Susana llegan a los asaltos, ambas tienen más o menos el mismo comportamiento promedio; solo difieren en la erraticidad con la que llegan, fenómeno que la ciencia de los pronósticos llama varianza. Varianza es dispersión, baja para Laura y alta para Susana; no inventemos palabras nuevas para llamar al agua tibia.

Y a fin de ser consecuentes con el catecismo predictivo de este libro, a la diferencia entre el horario promedio en que llegan y el pautado se la llama sesgo. En la jerga recién aprendida, Laura y Susana llegan, "en promedio", al mismo horario (el pautado para el ágape danzante, como diría mi tía Raimunda), o sea, son insesgadas. Pero el horario de llegada de Susana es mucho más disperso que el de Laura, es decir, tiene mayor varianza.

Sesgo y varianza (que, dichos de esta forma, parecen nombres de animadores de fiestas infantiles) son el pan y la manteca de la ciencia de la predicción. Porque, como quizás ustedes sospechen, predecir correctamente tiene que ver con que, en esa dupla, uno y otra sean pequeños.

Y avancemos más. A la suma del sesgo y la varianza se la conoce como error cuadrático medio. Bueno, sí, en realidad es la suma del sesgo al cuadrado y la varianza. ¿Por qué al cuadrado? Porque no estamos distinguiendo si el sesgo es por exceso o por defecto. En este caso, no nos importa más si alguien llega tarde que si llega temprano, solo tomamos en consideración si llega a destiempo. La

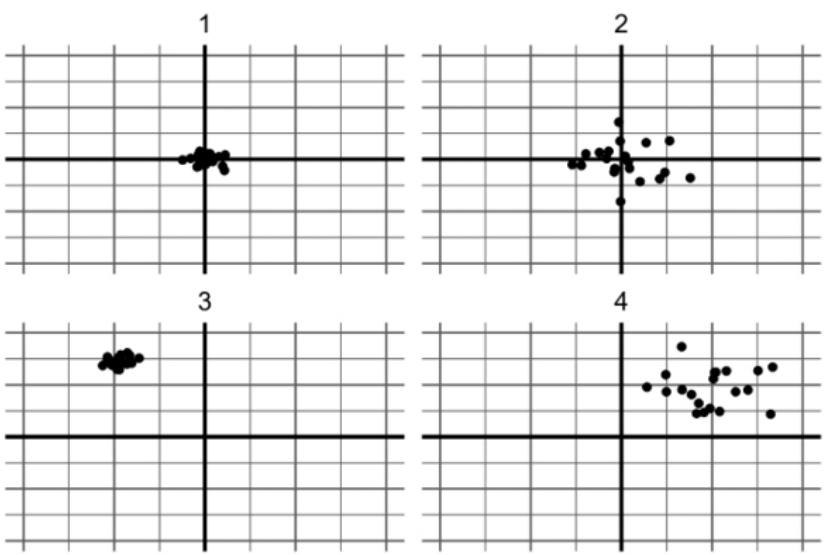
demora de diez minutos molesta tanto como la anticipación de diez minutos. Elevar al cuadrado es un viejo truco de la matemática para "hacer desaparecer el signo"; así, 10^2 es igual a $(-10)^2$.

El error cuadrático medio es crucial en la ciencia de los pronósticos. Es la medida más usada para juzgar pronósticos, obvio que no la única. Fíjense que da 0 solo si el sesgo y la varianza son 0, o sea, si un fenómeno se puede predecir con exactitud. Y también indica que predecir mal tiene que ver con que el sesgo y la varianza sean grandes. Esta idea es fundamental y amerita toda una sección aparte. Ahí vamos.

El hombre del rifle: sesgo, varianza y error cuadrático medio

A fin de entender esta cuestión del sesgo y la varianza, y cómo afectan al error cuadrático medio (que llamaremos ECM, para ahorrar tinta), volvamos al rifle del capítulo 4 cuyo objetivo es, naturalmente, dar en el blanco. Aquí debajo, la figura 5.1 muestra 4 "blancos" para el lanzamiento de cuatro rifles distintos, y 20 disparos en cada uno. Para facilitar la discusión, supondremos que siempre dispara la misma persona y en las mismas condiciones (viento, ruido de ambiente, etc.), porque lo que estamos estudiando es la performance de los rifles y no otra cosa.

Figura 5.1



El gráfico 1 de la figura 5.1 muestra los resultados de un rifle que

tiene sesgo bajo (tal vez nulo) y varianza también baja. El sesgo tiene que ver con la relación entre la posición promedio de los disparos y el objetivo (el blanco). En este caso, por lo menos visualmente, los disparos se distribuyen dejando el blanco en el centro. A diferencia de lo que ocurre en el gráfico 3, donde el promedio de los disparos está corrido hacia arriba y a la izquierda del blanco. En el primer caso, diremos que el rifle es insesgado ("centrado"), y sesgado ("corrido") en el caso del gráfico 3. Es a este corrimiento con respecto al objetivo al que llamamos sesgo.

Pensemos ahora qué sucede con la varianza, que, como dijimos, es un nombre sofisticado para la dispersión. Cuando pasamos del rifle del gráfico 1 al del 2, el centro de los disparos sigue siendo, aproximadamente, el blanco, pero ahora los disparos están más dispersos. En nuestro lenguaje, los rifles 1 y 2 son insesgados, pero el 2 tiene más varianza (es más disperso). Con relación al rifle 1, el número 3 es sesgado, pero igual de disperso. Finalmente, el rifle 4 parece ser el peor de todos: es disperso y está sesgado. En esta competencia, parece ganar el rifle 1: está centrado (es insesgado) y es preciso (tiene baja varianza).

Volvamos ahora a nuestras discusiones sobre predictibilidad. Todo parece sugerir que el ámbito más favorable para la predicción es el de sesgo y varianza bajos, que implican ECM bajo. Pero estos mismos ejemplos sugieren que, fuera de este contexto ideal, no es trivial si es más fácil lidiar con el problema del sesgo que con el de la varianza.

Estudiemos primero el problema del sesgo. De por sí, la palabra "sesgo" está socialmente "cargada": denota algo malo, tramposo. Así, cuando se habla de un periodista o crítico sesgado, se lo hace en un sentido peyorativo. Pero es importante ver

que el sesgo sea un problema es una cuestión de grado,

que el sesgo no es un problema severo si se tiene alguna idea de

cómo revertirlo.

Hace casi veinticinco años que visito a mi hermano Gabriel, que vive con su familia en Brasil; en la época en la que acostumbro a viajar, ese país tiene una hora de diferencia con la Argentina. Ajeno a los gustos sofisticados, uso relojes baratos, digitales y de plástico, de la marca que le disgusta a Shakira, de esos en los que, para cambiar la hora, hay que atravesar un mundo de botones minúsculos y menús hostiles a mi presbicia y mi impaciencia. Entonces, cuando vuelo a San Pablo, opto por dejar mi reloj así como está, e incorporar un tic mental que suma una hora a la que veo en su "display".

En el lenguaje de nuestras disquisiciones, mi reloj es igual de preciso en Buenos Aires y en San Pablo pero, una vez que aterrizo en Guarulhos, se vuelve sesgado: atrasa exactamente una hora. Entonces, ¿mi reloj deja de servir una vez que llego a Brasil? No, porque conozco el sesgo y sé cómo corregirlo, por lo menos mentalmente. Distinta sería la situación si, además de atrasar una hora, mi Casio se tornase también errático una vez en tierras de Pelé. En el caso realista, el sesgo de mi reloj no es un problema: se arregla sumando una hora.

Otro ejemplo en el cual los sesgos no resultan problemáticos es cuando, con objetivos de predicción, tolerar un pequeño sesgo implica una caída sustancial en la varianza, argumento que –recuérdenlo– será crucial en nuestras discusiones sobre IA y machine learning para la predicción, que abordaremos más adelante en este capítulo.

La naturaleza peyorativa del sesgo les hace creer a muchos que cualquier sesgo es inaceptable o, peor aún, que uno está necesariamente mejor cuando no los hay, ignorando por completo lo que sucede con la varianza. A tal efecto, hay dos chanzas de economistas que ilustran muy bien esta situación.

Empecemos por el chiste, que se refiere a dos economistas que van a cazar. El primero divisa un pato, toma su rifle, apunta, contiene la respiración, jala el gatillo y el disparo pasa un metro a la derecha

del pobre animal. Ahora es el turno del segundo economista, que toma el rifle de su amigo, apunta, y su disparo pasa también a un metro, pero a la izquierda del pato, que respira tranquilo. Los economistas se miran y dicen, contentos, "¡le dimos!". En nuestro lenguaje, el rifle que usaron es insesgado (en promedio, da en el blanco) pero errático, y en ningún disparo mata al pobre patito. Es decir, cualquier garantía de que un rifle sea insesgado es en promedio, o sea, no hace referencia a ningún disparo en particular sino al promedio de los disparos.

La segunda chanza que ilustra cuán inconducente es priorizar tercamente la "insesgadez" es la siguiente. Ofuscados por el resultado anterior, los economistas vuelven a la armería, a comprar un buen rifle. El vendedor les ofrece dos alternativas. El primero tiene la siguiente característica: el disparo sale, al azar, un metro a la derecha o un metro a la izquierda del blanco al cual uno apunta, más o menos como el rifle del ejemplo anterior. El segundo rifle dispara y siempre, y exactamente, el proyectil impacta a un centímetro a la derecha de donde uno apunta. Los economistas recuerdan su aprendizaje en lo más granado de la estadística clásica, donde les dijeron que los sesgos son pecados mortales, y eligen… el primer rifle.

El primero es insesgado (en promedio, da en el centro), pero, para beneplácito de los proteccionistas, con certeza nunca mata a nadie. El segundo rifle es sesgado, pero, a diferencia del anterior, tiene una dispersión nula. Peor aún, el sesgo es fácilmente corregible: es cuestión de apuntar un centímetro a la izquierda. Y matar, mata siempre (por lo menos, bichos voluminosos, no mosquitos ni piojos).

El partido de la predicción es de sesgo y varianza. No es exagerado pensar que la principal contribución de la IA, big data y machine learning es apelar a cuanto se pueda del combo datos-algoritmos para usar métodos centrados (con bajo sesgo) y precisos (con baja varianza). Un enorme paso adelante se dio cuando la estadística descubrió que, lejos de ser cuestiones aisladas, el sesgo y la varianza están atadas como un perro a su dueño. Más aún, ocurren en una solución de compromiso: para bajar el sesgo hay que tolerar más varianza, y viceversa. En otras palabras, esto del sesgo y la varianza

es un problema de “frazada corta”: al resolver un problema, se crea otro. Negociar este conflicto tal vez sea la parte más sutil y compleja de la actividad predictiva.

Pero antes de meternos en ese berenjenal, les adelanté que no es obvio que los errores por exceso importen tanto como los que ocurren por defecto: no es lo mismo llegar tarde que temprano, declarar culpable a un inocente o lo contrario, que sobren empanadas o que falten. Indagar las consecuencias de estas asimetrías es el tema del siguiente apartado.

¿Qué se puede hacer, salvo ver películas? Evaluación de errores de pronóstico

Mi amigo Ricardo es un fan del cine. Goza como loco con el séptimo arte, tiene una cultura cinematográfica envidiable y odia quedarse afuera de las discusiones sobre películas, actores y actrices. Con orgullo narra las tardes de su infancia en el cine de barrio (el "continuado" de tres películas por el precio de una) y de los bodrios infames que tuvo que fumarse para que no quedasen huecos en su inventario. Dolores es el extremo contrario: más allá de un par de películas a las que no pudo evitar llevar a su pequeña hija, el cine no es lo suyo. Disfruta tibiamente de lo que sus amigos tildan de clásicos y puede pasarse una semana entera reprochándose haber visto una película que juzgó como mala.

Esta situación ilustrará uno de los enormes límites (más bien, desafíos) que tiene la IA a la hora de hacer pronósticos: el diseño de un mecanismo predictivo no es independiente de cómo se lo evalúe. Es decir, la forma en la que se piensen ciertos pronósticos depende del usuario y su contexto.

Sí, lo sé: todo demasiado abstracto; vayamos a Ricardo y Dolores para ver cómo funciona la cosa. Supongamos que, entusiasmados con esto de la IA y big data, ambos deciden confiar en una app para que les sugiera una película. Esta actividad es la base fundamental de los sistemas de recomendación que mueven a la industria del entretenimiento, y una pieza clave de plataformas como Netflix o Amazon Prime.

En el contexto de las películas, hay dos errores posibles que el sistema de recomendación puede cometer:

recomendar una película que el usuario luego juzga como mala,

tildar de mala una película que evaluaría como buena.

En la jerga estadística, a uno de estos errores se lo llama “error de tipo I” y al otro, “error de tipo II”. A casi cuarenta años de la primera vez que tomé un curso de estadística, y tras haberle dedicado toda mi energía profesional a esta disciplina, todavía no logro recordar cuál es el de tipo I y cuál el de tipo II. También me cuesta distinguir entre cóncavo y convexo, o numerador y denominador. Debería hablarlo con mi analista.

Lo ideal sería no cometer errores. Pero la naturaleza azarosa de los gustos y las películas hace que sea inevitable. De hecho, no existe película (o canción o libro) exenta de disenso. Hay detractores de Tarantino, Coppola, Borges, Beethoven y los Beatles, y en el cable siguen pasando bodrios infumables que para algunos son películas de culto. Así, el arte demanda riesgo: todos leímos un libro que no nos gustó, o salimos decepcionados de un concierto o de una obra de teatro, y lo entendimos como un minúsculo precio a pagar por todas las veces que disfrutamos a lo loco.

Descartada la posibilidad de no cometer errores, resta la de pensar en un mecanismo predictivo que los minimice. Esta es una estrategia honesta y recurrente de la matemática aplicada: cuando un problema no tiene solución, intentar resolver uno relevante que sí la tenga. El que no tiene solución es el de predecir películas a riesgo nulo. Intentaremos entonces probar con el camino de “negociar” los errores, para que, dado que ocurrirán, molesten lo menos posible.

Y aquí viene el problema. La solución a esta cuestión dependerá de qué entendamos por errores y cómo los evaluemos. En este contexto, errores son dos: el de tipo I y el de tipo II que vimos más arriba.

Entonces, nos queda por resolver cómo sopesamos estos errores. El grueso de los cursos básicos de ciencia de datos apela a una estrategia pedagógica y computacionalmente conveniente: los trata

de manera simétrica. Es decir, creen que molesta de la misma manera perderse una gran película que comerse un bodrio. Esta lógica simétrica (los dos errores importan por igual) muchas veces es razonable. Por ejemplo, en tirar al blanco, lo que importa como error es cuán lejos está el disparo del blanco, no si a la derecha, a la izquierda, arriba o abajo. Esta percepción simétrica de los errores gobierna la construcción de la mayoría de los mecanismos predictivos.

Pero, lamentablemente, muchas veces la realidad es asimétrica. A modo de ejemplo, lo que les preocupa a las escuelas es que los alumnos lleguen tarde, y penalizan a quien llega un minuto después de que se cierra la puerta de entrada al colegio, a las 7.50 hs exactamente, como cuando yo fui a la primaria. Por otra parte, alguien que llegue a las 7.46 no es felicitado ni reprimido (más bien, resulta ignorado). Tal vez, llegar demasiado temprano resulte molesto para las autoridades, que a esa hora no tienen suficiente personal para cuidar a los chicos. En este contexto, un error importa mucho (llegar tarde) y el otro menos (llegar temprano). En este caso, la evaluación de los errores es asimétrica: como apuntaba antes, no es lo mismo llegar media hora tarde que media hora temprano.

En el caso de Ricardo y Dolores, lo que deben haber adivinado es que ellos tienen una percepción asimétrica del error que puede cometer un sistema de recomendaciones. A Ricardo le molesta mucho un error (perderse películas potencialmente buenas) y el otro no parece preocuparle demasiado (clavarse con un bodrio), y con Dolores pasa exactamente al revés. Entonces, un algoritmo de recomendaciones aparentemente justo, que sopesa ambos errores por igual, es rechazado tanto por Ricardo como por Dolores. Ricardo lo juzgará como timorato y Dolores como demasiado audaz. Ambos disconformes.

Entonces, un algoritmo de predicción de películas debería internalizar que a Ricardo le preocupa demasiado un tipo de error y poco el otro, y debería “retunearse” teniendo en cuenta esta premisa, y lo mismo para Dolores. En el extremo en el que Ricardo no tolerase en absoluto perderse películas, el algoritmo óptimo debería predecir “toda película es buena”, lo cual elimina por

completo el error que preocupa a Ricardo y comete errores que no le importan (ver bodrios). Para Dolores, misma cosa, pero al revés: si sus preferencias son tales que no puede tolerar en absoluto ver una película que no le gusta, el algoritmo tuneado para ella debería predecir que toda película es mala. Ricardo va siempre al cine; Dolores, nunca. Ambos contentos.

En la práctica moderna, este “retuneo” es un proceso automático, que surge de interactuar con los usuarios a la luz de sus experiencias. La idea de “learning” (aprendizaje) en machine learning tiene que ver con esta forma automática e iterativa de ir adaptando los algoritmos predictivos a la luz de la experiencia. Les habrá pasado que el algoritmo de Netflix o Mercado Libre prediga para ustedes cosas más afines a sus gustos a medida que pasa el tiempo e interactúan con la plataforma. Es IA en acción, el algoritmo autoadaptándose a más información sobre gustos y preferencias.

Vamos a explorar ahora un ejemplo más realista de tratamiento asimétrico de errores predictivos.

¿Es posible evitar la sobreventa en los vuelos?

Fue viral. En abril de 2017 las redes sociales estallaron con el espeluznante video en que se ve al médico vietnamita David Dao retirado por la fuerza de un avión de United Airlines. Se trataba de un vuelo sobrevendido y, hasta unos minutos antes del despegue, la aerolínea había ofrecido varias alternativas a quienes voluntariamente accediesen a postergar sus viajes y descender. Aun después de que la aerolínea anunciara una recompensa monetaria importante, la cabina seguía abarrotada, por lo que el comandante de a bordo anunció que, "de acuerdo con un algoritmo", cuatro personas habían sido elegidas al azar para abandonar el avión. Una de ellas era el doctor Dao, quien adujo que sus compromisos con sus pacientes le hacían imposible abandonar el vuelo, lo cual dio lugar a los episodios de violencia que se difundieron por todo el mundo.

Ahora bien, en tiempos de big data, machine learning e IA, ¿no es posible predecir exactamente quién vuela y quién no, de modo que no exista la sobreventa y no ocurran estas desagradables situaciones de las cuales muchos fuimos testigos (cuando no víctimas)?

El problema se relaciona con la asimetría en la valoración de los errores referida antes. Desde la perspectiva de las líneas aéreas, una vez que el avión despega, un asiento vacío es como el pescado del vendedor que baja de su barca y no tiene una heladera: al final de la jornada, es invendible, hay que tirarlo. La situación más común es que haya pasajeros que habiendo comprado su ticket no vuelan, y que casi siempre haya pasajeros en lista de espera, a la expectativa de que se libere un asiento. Por supuesto, volar con asientos vacíos implica una pérdida potencial para las aerolíneas, en términos de los ingresos perdidos por no haber vendido esos vuelos a los pasajeros en la lista de espera. La sobreventa tiene otros costos, en términos de los conflictos operativos y reputacionales que causa dejar a pasajeros sin viajar, aun cuando habían comprado un pasaje y confirmado el vuelo, como el caso del doctor Dao.

Sin lugar a duda, las empresas aéreas recurren a poderosos algoritmos que usan todo tipo de información (desde el género hasta si uno tiene preferencias alimentarias), para predecir las chances de que un pasajero no suba al avión aunque haya comprado su ticket. Asimismo, otros algoritmos predicen las chances de que haya pasajeros en listas de espera para cada vuelo. De este complejo proceso surge una predicción de la cantidad de asientos vendidos para ocupar un vuelo que, IA y big data mediante, es bastante precisa.

Ahora bien, la empresa sopesa los dos errores posibles:

salir con asientos vacíos,

salir con el avión lleno y absorber el costo operativo y de marketing de que se enojen los pasajeros que no podrán viajar a pesar de haber pagado su pasaje.

La línea aérea saca cuentas y llega a la conclusión de que estos errores no son simétricos, y que, en general, le es más fácil (económica y comunicacionalmente) lidiar con el enojo de los pasajeros varados. Entonces, el algoritmo es “tuneado” para impedir el primer error (que el avión salga con asientos vacíos) a costa del segundo: la sobreventa. Las empresas del rubro argumentan que, de no existir la sobreventa, ellas tendrían que absorber el costo de volar con asientos vacíos, si el algoritmo tratase los errores de forma simétrica (esto redundaría en pasajes más caros y menos alternativas de vuelo).

En todo caso, esta historia ilustra que la elección de un método predictivo está guiada por lo que se entienda por error y cómo se lo evalúe. Si para las aerolíneas bajar a la fuerza pasajeros del avión representase un perjuicio simétrico al de volar con asientos vacíos, un algoritmo “negociaría” entre estos dos errores. Entonces, que

haya sistemáticamente sobreventas no es una falla algorítmica, sino la consecuencia elemental de que a la empresa no le da lo mismo un error que el otro.

Este problema de asimetría es más frecuente de lo que uno sospecha. En el ámbito bancario, una predicción importantísima es si un cliente pagará o no un crédito. Por cierto, al banco no le da lo mismo otorgar un crédito a un cliente que no va a pagarlo que negárselo a alguien que eventualmente lo pagaría. La asimetría entre estos errores es una parte fundamental de la gestión de riesgo de un banco, y la forma de tratar estos errores es una tarea delicada. En el derecho, esta asimetría es crucial: al sistema judicial no le da lo mismo tratar a un inocente como culpable que a un culpable como inocente. La Declaración Universal de los Derechos Humanos señala en su art. 11 que toda persona acusada de delito tiene derecho a que se presuma su inocencia mientras no se pruebe su culpabilidad, lo cual implica una marcada asimetría en el tratamiento de los errores de juicio.

Si alguna vez entraron a la página de una empresa de inversiones (o a la parte correspondiente de un banco), verán que el primer paso es completar un extenso cuestionario acerca del “perfil de inversor”, que lo que hace es, justamente, ver cómo uno evalúa los riesgos subyacentes en una inversión.

No entender cómo sopesar errores es, tal vez, uno de los principales (perdón) errores que cometen quienes se inician en la ciencia de datos. Ningún algoritmo sensato elimina errores por completo, sino que los negocia según algún criterio. Lo importante es que este criterio no es una cuestión meramente algorítmica: depende de las preferencias de quien se beneficia o perjudica de los resultados del modelo predictivo. Dicho de otra forma, el algoritmo no tiene ninguna forma obvia de saber de antemano las preferencias de quien lo construye. Es la interacción con el cliente (la línea aérea, por ejemplo) lo que le permite aprender al analista (y al algoritmo) que, en este caso, es preferible un tratamiento asimétrico de los errores. ¿No puede el algoritmo eliminar la sobreventa? El algoritmo puede; pasa que la aerolínea no quiere. Acá la culpa no es del chancho, sino de quien le da de comer.

¡Lo fermizamo' con alambre! El rol de los algoritmos

En cuanto a la materia primordial de este libro, las predicciones, ustedes notarán que en repetidas ocasiones diremos que los avances en la performance de muchos métodos tienen que ver con poder simular situaciones con una computadora. Y ha llegado el momento de ocuparnos de qué significa esta cuestión, en la cual parece que estamos reemplazando análisis por fuerza bruta. Para adentrarnos en el universo de las simulaciones, déjenme empezar con un ejemplo que, más o menos, todos conocemos.

"¿Cuántos paquetes de 5 figuritas habría que comprar para poder llenar un álbum de 200 figuritas?", preguntó alguien en Twitter. La gran mayoría de sus colegas intentó eludir la pregunta ("¿Para qué querés juntar figuritas?"), cambiar el contexto ("Te conviene ir al Parque Rivadavia a cambiarlas con otros coleccionistas"), o repreguntar ("¿Hay difíciles?", "¿hay repetidas?"). Obviamente, muchos se limitaron a un escueto "depende", la respuesta clásica del economista arrinconado contra las sogas. Pero muy pocos aventuraron una respuesta.

Y no cabe duda: enfrentar de forma apropiada la pregunta en cuestión depende de cierta contextualización. El más simple de los contextos presupone que las figuritas se distribuyen aleatoriamente en los paquetes –de modo que no hay figuritas difíciles– y que la pregunta se refiere a cuántos paquetes se espera comprar, ya que la propia aleatoriedad de la situación impide dar una respuesta certera. En esta circunstancia (irrealista, por cierto), la respuesta correcta es 233 paquetes, y hay dos maneras de obtenerla. Una es por medio de un cálculo matemático, sofisticado para el lego en cuestiones numéricas, pero simple (aunque desafiante) para quien posee una mínima formación universitaria en probabilidad.

Una vía alternativa es echar mano a una simulación computacional, es decir, reproducir muchas veces en una computadora el azar de comprar figuritas y ver cuántos paquetes hizo falta comprar, en

promedio, para llenar el álbum en cuestión, estrategia que demanda mínimos conocimientos de programación.

La respuesta analítica a la versión simple del problema apareció de manos de un físico y de un estudiante doctoral de economía en el exterior. La proveniente de la vía computacional llegó unos minutos después, con un joven programador que, además, compartió en Twitter las pocas líneas de código que permiten computar la respuesta con asombroso grado de precisión.

En circunstancias más generales que las del ejemplo de las figuritas, la aproximación algorítmica requiere pensar con detalle cómo armar una "máquina computacional" que funcione como el problema estudiado. En el ejemplo mencionado, el primer paso consiste en precisar la cuestión, en reaccionar de manera creativa ante la indefinición de una pregunta genérica como la formulada. La realidad no hace preguntas como las de un examen universitario, sino que las plantea difusamente, y es trabajo del analista reformularlas sin que se pierda relevancia. Los que proveyeron una respuesta lo hicieron fijando un contexto (distribución aleatoria de figuritas en paquetes, inhabilidad de cambiar figuritas en un mercado, etc.). Este proceso de contextualización es un primordial primer paso en la especificación de un problema, y el abecé de un proceso algorítmico.

Una elocuente muestra de la importancia de fijar contextos refiere a una clásica pregunta que hizo el físico Enrico Fermi a sus alumnos, en la primera mitad del siglo XX: "¿Cuántos afinadores de piano hay en Chicago?". Por cierto, no existe ninguna forma indiscutible ni simple de abordar semejante pregunta. Fermi comienza suponiendo que, en esa época, en Chicago,

había unos 5 millones de habitantes,

aproximadamente 2 personas por hogar,

de cada 20 de esos hogares, 1 tenía un piano,

ese piano era afinado una vez por año y

afinar un piano toma unas dos horas de trabajo.

Con esos números, Fermi estimó que había 125 afinadores.

Lo importante de este ejemplo no es la respuesta, sino la explicitación de una serie de "subpreguntas" (y sus respectivas respuestas), que conducen al resultado final, dividiendo el problema en porciones que pueden ser escrutadas cada una por separado. Eso le permite al analista reconsiderar sus cálculos si luego dispone de más tiempo o recursos. El instinto de partir problemas grandes en pequeños es central a lo algorítmico y al razonamiento sistemático en general. En el ejemplo de las figuritas, como dijimos, el contexto simple implica suponer que están distribuidas al azar en los paquetes y que no hay un mercado para el intercambio, lo que permite abordar el problema con unas pocas líneas de código de programación. Esas líneas pueden ser alteradas para acomodar la existencia de figuritas difíciles o, con más esfuerzo, de intercambios.

Lo que esta historia ilustra es que, muchas veces, una simple estrategia de simulación permite aproximar con asombrosa precisión problemas predictivos de difícil (cuando no imposible) abordaje con herramientas matemáticas clásicas, como el cálculo o el álgebra. El rol de las simulaciones en varios aspectos científicos es doble. En primer lugar, como vimos, las simulaciones permiten llegar a soluciones en contextos en que los métodos tradicionales no están disponibles o son muy costosos. Pero un aspecto importante de la simulación es que la naturaleza secuencial de un algoritmo demanda explicitar y separar supuestos, datos y procedimientos, lo cual redunda en una elegancia de razonamiento que, a su vez,

facilita la detección de errores y la reformulación de estrategias para resolver problemas. La lógica de la programación va mucho más allá del uso de un lenguaje específico y remite a formas de razonar atemporales. Vayamos entonces a explorar por qué una simulación puede aproximar soluciones a varios problemas analíticos.

Otra vez arroz: el rol de las simulaciones

En el capítulo 1 contamos que había dos formas de argumentar que las chances de que al tirar una moneda salga cara son del 0,5. Una es meterse en el universo de la física y la anatomía, y ver que no hay razones para predecir que conviene apostar a una cara más que a la otra. La segunda ruta es lanzar la moneda repetidas veces y aplicar la LGN. Una simulación es una versión "con esteroides" de esta segunda ruta. Y aquí conviene que nos detengamos con más detalle en cómo funciona esta alternativa que, mucho más allá de las moneditas, ocupa un espacio fundamental en los pronósticos modernos de la meteorología, las finanzas o los deportes.

Una parte no menor de mis clases fluye en un fárrago de ecuaciones. Como no me gusta monologar, cada tanto paro en el medio de un razonamiento matemático e indago a los alumnos (de hecho, apelo recurrentemente a ese yeite en este libro). Por ejemplo, antes de terminar de demostrar un resultado, pregunto: "Y esto, ¿cuánto creen que da?" o "Si seguimos con este procedimiento, ¿cuál es el valor que tiende a tomar tal o cual variable?", y otras preguntas por el estilo, con la idea de que los alumnos tengan una disposición activa y deduzcan ellos mismos "quién es el asesino", porque, a la larga, la matemática es tal vez la más hermosa de las novelas de misterio.

Una vez noté que un alumno respondía con convicción a mis preguntas, y acertaba con precisión algunas y erraba por completo otras. Raro. Por lo general, los buenos alumnos aciertan siempre (o responden solo cuando están muy seguros), y el resto no comete errores porque no participa. Luego de un tiempo y de repetidos intercambios con este alumno, le dije: "Me llama la atención que tus respuestas sean a veces hiperprecisas, y otras, disparates". El pibe me respondió: "Aaah, es fácil. Me di cuenta de que casi todo lo que preguntás tiene como respuesta cero, uno o infinito; entonces, cada vez que preguntás cuánto da esto o a dónde tiende lo otro, yo digo

algunas de las tres posibilidades, al azar, y medio que a un tercio de tus preguntas le pego". Genio total. A partir de esta experiencia empecé a ser más cuidadoso con mis preguntas (ni hablar de que más que el resultado me importa saber de dónde lo sacaron).

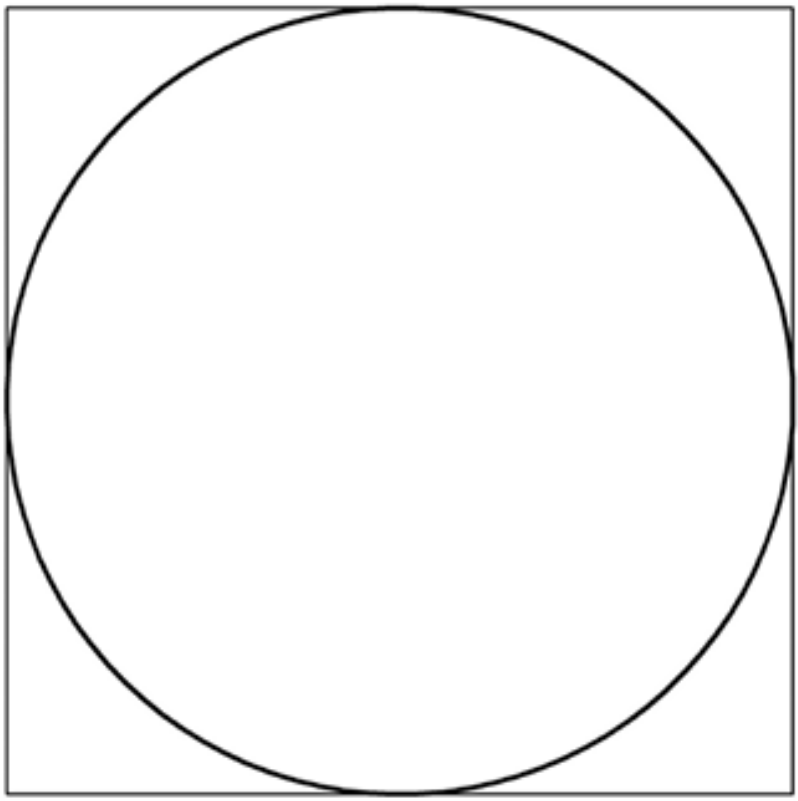
Pero lo interesante de esta anécdota es el instinto de este alumno, de involucrar el azar para un problema que no tiene nada de aleatorio. Por ejemplo, el área debajo de cualquier curva de densidad da 1. No importa si no entendés de qué cuernos estoy hablando, que dé "1" es una propiedad de las curvas de densidad: nada azaroso en la pregunta ni en la respuesta. De haber hecho en clase la pregunta: "¿Cuánto da el área debajo de esta curva de densidad?", mi alumno azaroso, tras sacarme la ficha, habría respondido, al tuntún, que 1, que 0 o que ∞. Si se trata de entender, es obvio que no ganó nada, pero si el desafío es dar una respuesta acertada a mis preguntas, le va mejor que a quien nunca dice nada; son dos cosas distintas. Y, como buen tema recurrente, reaparece en este libro sobre pronósticos una aclaración: entender es una cosa; predecir, otra; tienen mucho en común, pero no son lo mismo.

Chanzas apartes, una enorme contribución de la computación moderna consiste justo en involucrar al azar (como mi alumno) para problemas que no tienen nada de aleatorio y así abrir una ruta para enfrentar cuestiones con las que es muy complicado lidiar con métodos tradicionales.

He aquí un caso simple que ilustra este tipo de situaciones. El área de un cuadrado es la longitud de cualquier lado (son todos iguales) elevado al cuadrado. Así, el área de un cuadrado de 20 cm de lado es 20 cm al cuadrado, vale decir, 400 cm^2, como les enseñaron en la escuela y seguro se acuerdan. Simple. ¿Hay azar en alguna parte? No, mi capitán. Bien, sigamos.

¿Y si ahora queremos calcular el área de un círculo inscripto en este cuadrado? Digo, de un círculo que tiene la forma de la figura de aquí abajo, la 5.2:

Figura 5.2



Uff… es más difícil. Tiene que haber una formulita, pero, no sé ustedes, yo me la olvidé (si es que alguna vez la supe). Calculo que mis conocimientos de matemática me permitirían deducirla rápido; pero como me da mucha pereza, se la pregunté a Roccatagliata, que, listo como un bombero, me la dijo al toque: es π por la superficie del cuadrado, dividido 4. OK, procedamos:

$$\frac{3{,}1416 \times 20\text{ cm} \times 20\text{ cm}}{4},$$

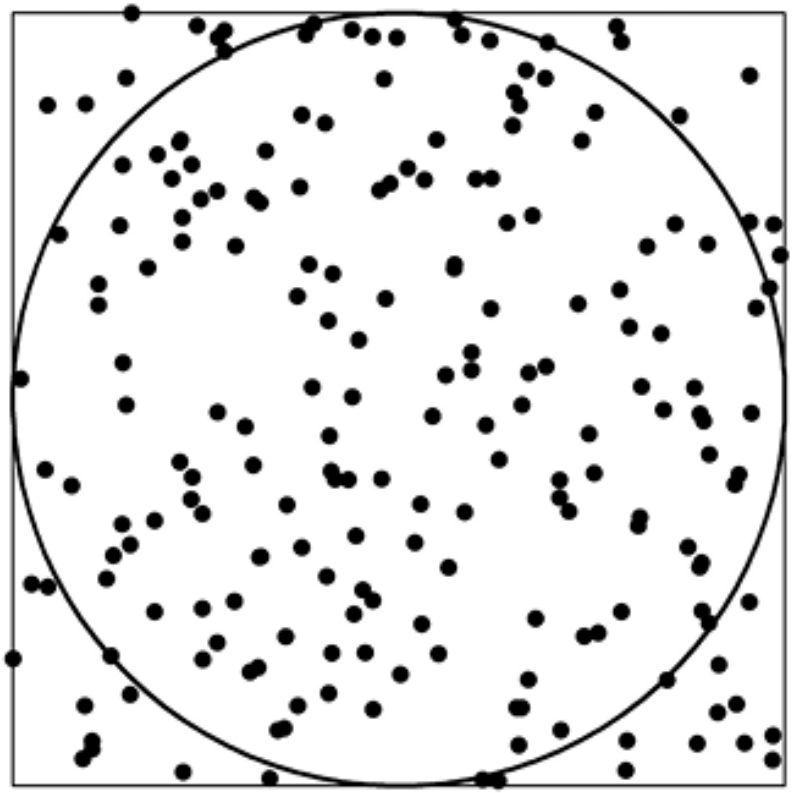
que da 314,16 cm^2, previsiblemente menos que el área del cuadrado. ¿Hay azar en alguna parte? No, mi capitán. Bien, sigamos.

Ahora apelemos a los instintos de mi alumno azaroso y exploremos una ruta alternativa para calcular el área del círculo en el cuadrado. Para hacerlo fácil, supongamos que tenemos una plancha cuadrada (yo tengo una, para hacer bifes) que mide 20 cm de lado. Con un marcador, dibujamos el círculo inscripto, como en la figura de más abajo, la 5.3. Ahora buscamos una bolsa de arroz, de la cual vamos a usar 200 granos. Luego revoleamos los granos de arroz de modo que caigan de manera uniforme en la superficie de la plancha. Parece una receta de cocina, pero no lo es. O sí, hay una fuerte relación entre los algoritmos y las recetas.

Y acá viene la magia: la superficie del círculo es, aproximadamente, la superficie de la plancha multiplicada por la proporción de granos que cayeron dentro en el círculo. ¿No me creen? Háganlo. ¿Sospechas? Háganlo varias veces. Es más, pueden usar este truco para animar fiestas infantiles, bautismos y despedidas de soltero; en este país, el que no es rico es porque no quiere. Ahh…, casi me olvido de "la" pregunta: ¿hay azar en alguna parte? Sí, mi capitán. ¿Cuándo? Cuando revoleamos el arroz. De hecho, si repetimos el experimento, va a dar muy parecido, pero no exactamente igual.

¿Qué es lo que ha pasado? Hemos reemplazado un problema analítico (encontrar el área), para el cual hay una fórmula y que, por ende, no tiene nada de aleatorio, por otro que involucra al azar (contar granos). ¿Qué ganamos? Y, que no tuvimos que pensar la fórmula del área de un círculo inscripto en un cuadrado. ¿Qué perdimos? Precisión. La respuesta que obtuvimos es una aproximación a la verdadera respuesta. No es difícil pensar que la respuesta será tanto más precisa cuanto más chicos sean los granos de arroz y más granos tiremos. ¿No me creen? Háganlo con semillas de sésamo y fíjense qué pasa. Cuando lo hice con la compu usando 200 puntos (y me dio el gráfico que muestro como figura 5.3), la superficie estimada era de 322 cm^2. Cuando lo hice con 10.000 puntitos, me dio 314,2 cm^2. ¡Nada mal!

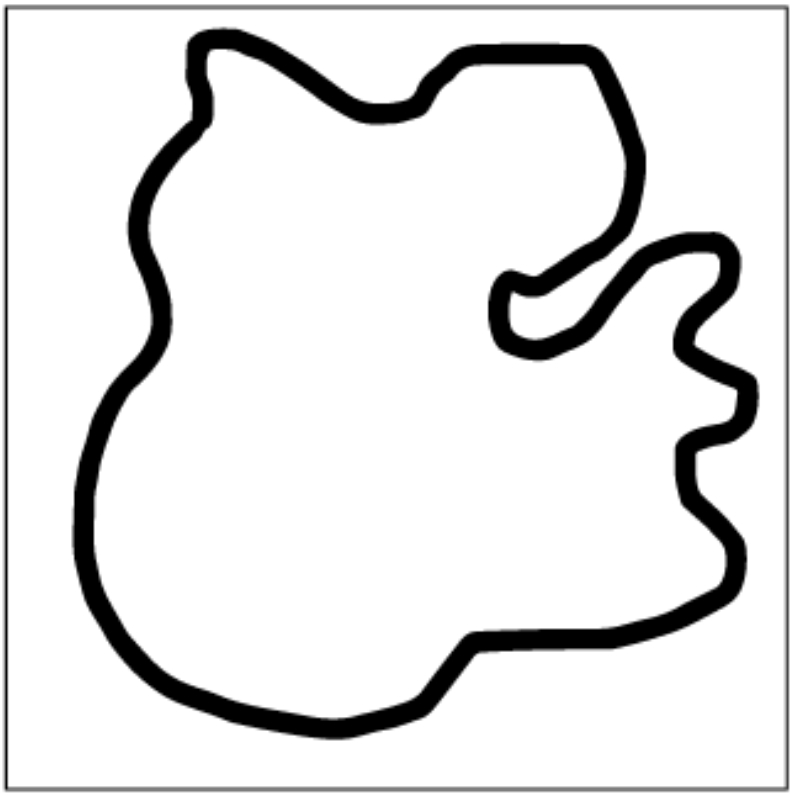
Figura 5.3



Revoleando globos terráqueos en Montecarlo

Convengamos en que este es un ejemplo demasiado simple, y que, máxime en tiempos de ChatGPT, los costos de dar con la formulita y sacar la cuenta son ínfimos en relación con los de conseguir una plancha, arroz, tirar los granos, contarlos, etc. Pero ¿qué pasa si ahora, en vez de querer calcular el área de un prolijo círculo, queremos la de la siguiente figura, la 5.4?

Figura 5.4



Y ahora los quiero ver. ¡Vaya a saberse cuál es la fórmula que gobierna a esta forma caprichosa! Así y todo, el método del arroz funciona exactamente de la misma manera: el área de esta forma es la proporción de granos que caen dentro de ella, multiplicada por el área de la plancha. Roccatagliata, a sacar arroz del medio. Acá los costos de buscar la formulita son enormes, si es que existe tal fórmula. Y los de implementar el método del arroz son casi los mismos.

Y les cuento una más, que aplicó en el aula la matemática argentina Daniela Parada, una auténtica entusiasta de la enseñanza de alta calidad. Este ejemplo está inspirado en el brillante libro de Richard McElreath (Repensemos la estadística puede ser una versión: en el original, no traducido, leemos Statistical Rethinking) que pronostico que revolucionará la enseñanza de la estadística. Supongamos que queremos estimar qué proporción del planeta Tierra está cubierta

por agua. Suena a muy difícil. La parte ocupada por el agua no parece tener ninguna forma amigable, como el círculo adentro de la plancha. He aquí la idea, basada en simulaciones. Tomemos un globo terráqueo, revoleémoslo en el aire, atrapémoslo con las dos manos y, cual batalla naval, anotemos si el dedo índice de la mano derecha cae en "agua" o "tierra". Repitamos este procedimiento muchas veces. Calculo que ya se dieron cuenta de a dónde vamos: la proporción del planeta Tierra ocupada por agua es la proporción de veces que en el procedimiento anterior salió agua. Daniela implementó recientemente este ejemplo en un congreso de enseñanza de la matemática. Fue conmovedor ver a los profesores revoleando el globo y anotando resultados, y les juro que idéntico resultado se obtiene en el aula, con chicos del secundario. Sí, hay mucho por trabajar en la enseñanza de la ciencia de datos… pero ese es otro libro.

La gran contribución de las simulaciones es que en ámbitos donde la complejidad manda –en meteorología, como vimos, pero también en finanzas, en economía y, como analizaremos en el próximo capítulo, en los pronósticos deportivos–, la aproximación computacional es casi la única vía posible para hacer pronósticos confiables.

Al método de usar el azar en vez de una fórmula analítica para calcular un área se lo llama método de Montecarlo o, para mayor precisión, integración por Montecarlo. Matemáticamente no es la gran cosa (en realidad, sí), pero conceptualmente es un salto cuántico: hemos involucrado al azar para lidiar con algo complejo pero que no tiene nada aleatorio, como mi alumno, como el que calculó la cantidad de paquetes de figuritas jugando con una simulación. No es exagerado decir que casi todos los experimentos y las simulaciones computacionales en estadística son, en espíritu, como el problema del cuadrado, el círculo y el arroz.

El inventor del método, el matemático polaco Stanisław Ulam, dice que se le ocurrió una vez que, estando enfermo, quería calcular ciertas probabilidades relacionadas con jugar al solitario. Sin la posibilidad de acceder a lápiz y papel, se dio cuenta de que podía conocer estas probabilidades al jugar repetidas veces al solitario y ver qué pasaba.

El nombre del método (que refiere al famoso casino de Mónaco) surgió en un contexto menos simpático. Ulam fue uno de los científicos involucrados en la construcción y lanzamiento de la bomba atómica que acabó con Nagasaki e Hiroshima en la Segunda Guerra Mundial, para lo cual había que realizar ciertos cálculos hipercomplejos que se beneficiaron de esta nueva técnica que apela al azar. En épocas de guerra y secreto, el método requería un nombre críptico (propio de una buena película de espías), y en ese contexto el matemático griego Nicholas Metropolis sugirió Montecarlo. El no menos cinematográfico nombre (bueno, apellido) Metropolis es fundamental en la cuestión predictiva. El algoritmo de Metropolis-Hastings es un componente clave de una idea que revolucionaría por completo la estadística y la computación en la década del ochenta del siglo anterior, el "método de Montecarlo basado en cadenas de Markov": para muchos, una de las tres ideas más revolucionarias de la ciencia de datos de los últimos cien años, y, sin exagerar, un componente fundamental de la revolución de machine learning y la IA.

Así las cosas, ¿la cuestión es dejar de aprender análisis matemático y reemplazarlo por simulaciones? De ninguna manera: de hecho, el método de Montecarlo tal como lo describimos involucra varios pasos analíticos. Uno de ellos, en nuestro ejemplo, es saber si un grano de arroz cayó dentro de la figura o no. En este ejemplito simple, es fácil hacerlo a ojo, pero en situaciones prácticas puede ser difícil enseñarle a una computadora esta tarea. A veces, dar con una fórmula que permita evaluar de forma automática si el grano de arroz cae o no en la figura es igual o más difícil que dar con la fórmula para computar su área, lo cual relativiza las ventajas de involucrar al azar. No es análisis versus simulación, sino análisis y simulación. No es computadora en vez de lápiz, son ambos.

Volver del futuro

Criticar a un pronosticador porque yerra es como cuestionar a un perro porque ladra. La forma correcta de juzgar a un pronosticador es viendo no si yerra, sino cómo lo hace. Apuntemos algunas ideas clave de este capítulo:

El error cuadrático medio (ECM) es la forma más popular de medir cuán grande es el error de pronóstico agregado, y se compone del sesgo (cuán lejos está el promedio de los pronósticos de lo que uno quiere predecir) más la varianza (cuán errático es el pronóstico).

Predecir bien consiste en que estos factores sean lo más pequeños posible. Manejar las “perillas” del sesgo y la varianza nos posiciona en el corazón mismo de la estadística y la IA moderna.

En muchas ocasiones cabe un tratamiento simétrico de los errores: juzgar como idénticamente malos los errores por exceso y aquellos por defecto.

Pero muchas circunstancias demandan un manejo asimétrico: no es lo mismo llegar tarde que temprano, dar un crédito que no se paga que negarse a dar uno que se pagaría.

Los bancos son “dadivosos” con las tarjetas de crédito, porque prefieren asumir el riesgo de que alguien no pague frente a perderse

la posibilidad de cobrar intereses a quienes sí pagarán sus deudas. Es decir, el primer error les importa (asimétricamente) menos que el segundo.

Las computadoras cumplen un rol fundamental en simular situaciones azarosas.

Simular es, en realidad, una manera inteligente de aprovechar el azar para aproximar algunas situaciones puramente analíticas, que, por su complejidad, a veces es mejor relegar a las computadoras que al razonamiento analítico.

Además de permitir resolver complejos problemas predictivos, el razonamiento algorítmico es central para enfrentar, planificar y evaluar procedimientos que lidian con lo complejo e incierto.

6. Complejidad, meteorología y fútbol

Parece estar de moda llamar “saber” al azar y, así, ganadores del Quini 6 pretenden venderse como “hábiles seleccionadores de números” y ofrecen charlas motivacionales para empresas. Escondidos detrás de las “virtudes” del azar, hemos sido testigos de las habilidades predictivas del pulpo Paul o, más recientemente, del cerdo Mystic Marcus, que, luego de una irrefutable sucesión de éxitos, casi termina al horno y con una manzana en la boca luego de predecir incorrectamente que Inglaterra le ganaría a Croacia en el Mundial de Fútbol de 2018.

La credibilidad predictiva de estos pobres bichos remite a una vieja chanza de las finanzas, que he contado en mis anteriores libros. ¿Cómo convertirse en un certero predictor de la Bolsa, sin gimnasia, sin pastillas, sin datos, sin algoritmos? Envíen 1000 correos electrónicos, en la mitad digan que el precio de una acción sube, en la otra, que baja. Al día siguiente, al grupo que recibió la predicción correcta envíenle 500 correos repitiendo la estrategia de a una mitad decirle que sube y a la otra, que baja. Al cabo de una semana, habrá unas 30 personas que dirán: “¡Hay un tipo que hace seis días le pega a la Bolsa!”. Así es como todos recordamos al pulpo Paul, algunos menos al cerdo Mystic Marcus, y casi nadie al puercoespín León, el hipopótamo Petty, el conejillo de Indias Jimmy, la cotorra Mani, el chimpancé Pino o al potamoquero Apelsin, y otros pobres bichos que, luego de algún acierto predictivo y sus consabidos quince minutos de fama, fueron condenados al olvido.

En este capítulo aventurero nos meteremos con dos cuestiones predictivas centrales para el comportamiento humano, que afectan a todos por igual: el tiempo y el deporte. En los capítulos iniciales nos focalizamos en la omnipresencia del azar como culpable de la naturaleza impredecible de muchos fenómenos. Aquí agregaremos un ingrediente clave: la complejidad que surge de lo no lineal y lo dinámico. Es la mezcla de estos factores lo que se esconde detrás de

la aparente impredecibilidad de las lluvias y de quién ganará un partido de fútbol.

Veremos que, más allá del folklore y la anécdota, el tiempo y el fútbol tienen varios aspectos predecibles, y nos meteremos en la trastienda de quienes llevan a cabo estas tareas con métodos científicos.

Mariposa technicolor: caos, no linealidad y condiciones iniciales

En relación con lo complejo, muchas veces preferimos una cómoda verosimilitud a una verdad urticante. Y en esa categoría cae la historia (que ya apareció en el capítulo 2, cuando hablamos de sesgos en encuestas) del famoso "efecto mariposa", indispensable para este libro sobre predicciones: todos lo conocemos, muchos lo repetimos como loros, unos pocos lo entienden. Comenzaremos por la versión más simple de todas, para luego escalar a una versión más compleja, que nos permitirá ver con precisión por qué ciertos fenómenos son difíciles de predecir.

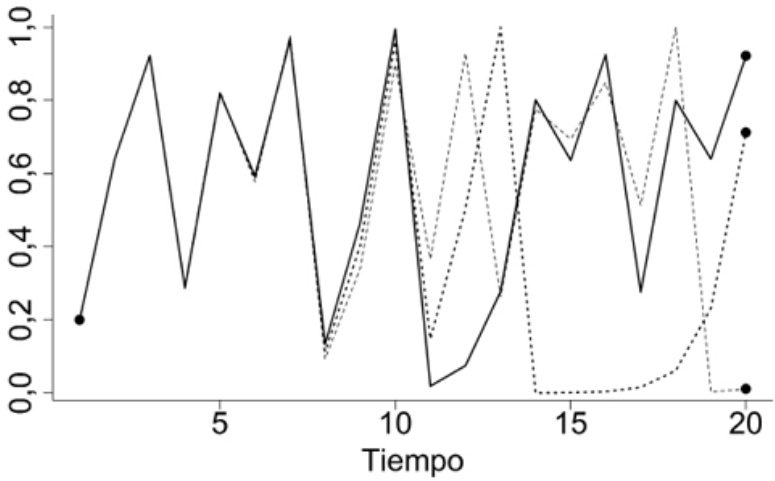
Vamos a la versión simple, la que todos conocemos. La frase icónica es: "El aleteo de una mariposa en Brasil puede producir un tornado en Texas". La anécdota no cambia si reemplazamos Brasil y Texas por otros dos lugares distantes. La frase refiere a la complejidad de ciertos fenómenos, en los que variaciones ínfimas en sus condiciones iniciales producen resultados aparentemente disparatados (caóticos) y, por consiguiente, impredecibles. En cuanto a la paternidad, la mayoría de las fuentes consultadas la atribuyen al matemático y meteorólogo Edward Lorenz, figura central en la teoría del caos, que versa sobre la dinámica de este tipo de efectos.

Pero a fin de entender cuál es la naturaleza de los fenómenos (im)predecibles, la anécdota del efecto mariposa resulta mucho más rica que lo que su versión "folklórica" sugiere, y vale la pena indagar más a fondo. La versión un poquito más complicada es que se trata de sistemas "apenas no lineales", muy sensibles a sus puntos de partida. Ahí vamos, no se asusten.

A continuación, les muestro un gráfico –la figura 6.1– en el que seguimos el comportamiento de tres series en el tiempo. Como ven, empiezan muy juntas; al ojo humano, diríamos que en el mismo lugar. Luego siguen más o menos unidas, pero a medida que

transcurre el tiempo su comportamiento comienza a ser evidentemente disímil, a punto tal que, hacia el final del gráfico, cada una anda por su lado, como los exintegrantes de cualquier banda icónica de rock, que luego de las consabidas “diferencias musicales” pasan de hermanos inseparables a enemigos públicos número uno.

Figura 6.1



El comportamiento errático de cualquiera de las tres series no parece corresponder al de las curvas que vimos en el secundario: la recta, la parábola, la exponencial, y ni siquiera al de las más aterradoras pero elegantes funciones trigonométricas, como el seno o el coseno. Por el contrario, la imaginación lleva a creer que detrás de estas series ingobernables están las fórmulas kilométricas de los pizarrones de Sheldon Cooper en The Big Bang Theory, sazonadas con una apropiada dosis de aleatoriedad.

Y acá viene la magia de la matemática. Las tres series que les mostré obedecen en realidad a una única fórmula:

$xt = 4xt\text{-}1 - 4x^2t\text{-}1$

Sí, esta aparentemente inocente formulita es una bestia capaz de engendrar los comportamientos caóticos del dibujo. Simple pero muy interesante. Los invito a explorar con más detalle cómo funciona este endriago. La notación “xt” hace referencia a la serie “x” en el período “t”. Si usamos los números naturales para denotar el tiempo (t = 0, 1, 2, 3, …), x5 es el valor de la serie x en el período 5. Usaremos x0 para referir al valor que toma x en el inicio, la condición inicial, protagonista clave de este thriller. La notación “xt-1” refiere al valor de x un período antes de t.

Hay varias cosas por discutir sobre esta regla, ecuación o fórmula. En primer lugar, es una fórmula dinámica, más concretamente, recursiva: el valor de la serie en un período depende de los valores que tomó antes. La recursividad tiene que ver con que la fórmula se realimenta: por ejemplo, el cuarto valor depende del tercero, pero, por la misma regla, el cuarto determina el quinto, y así sucesivamente. En otros términos, la formula “se muerde la cola”.

Ahora nos resta explorar el otro ingrediente diabólico, su no linealidad, que se relaciona con que, en la fórmula, aparece el pasado de x y el pasado al cuadrado; vale decir, x^2. Es este último término el que rompe la linealidad y genera lío. En síntesis, nuestra aparentemente inocente fórmula es dinámica y no lineal.

Y ahora ya tenemos todos los ingredientes para jugar con nuestra máquina caótica. Como les dije, las tres series fueron creadas con la misma fórmula, pero con tres puntos de partida apenas distintos: 0,1998, 0,2 y 0,2002. Estas son las famosas condiciones iniciales, los valores de x0 que, en cada caso, usamos para que el sistema arranque.

Ya con más lenguaje para caracterizar su comportamiento, volvamos a la figura 6.1. Las tres series empiezan con diferencias ínfimas (“el aleteo de una mariposa”), tanto que, en el gráfico, parece que tienen el mismo punto inicial. Al principio parecen

moverse en sincronía, pero a medida que pasa el tiempo, la evolución de cada una de ellas es muy errática y desincronizada (“un tornado en Texas”).

Este tipo de comportamiento caótico es muy engañoso, es un monstruo disfrazado de una inocente abuelita. Fíjense que, si supiésemos con exactitud el punto de partida y que todo el movimiento de la serie está gobernado por la fórmula descripta, podríamos predecir perfectamente todo lo que va a ocurrir con la serie, en cualquier momento, en un futuro tan lejano como quisiésemos. Ahora les propongo que nos concentremos en cualquiera de las tres series. Su comportamiento luce errático, caprichoso, caótico, pero, para nuestra sorpresa, la fórmula no incluye ningún elemento aleatorio. Esta es la “trampa” de los sistemas no lineales: ciertas configuraciones simples se “camuflan” muy bien con comportamientos aparentemente guiados por el azar.

En segundo lugar, si bien las tres series obedecen a la misma fórmula, su comportamiento difiere radicalmente de acuerdo con cómo fijemos los puntos de partida. Es decir, este minisistema caótico es hipersensible a las condiciones iniciales, y de ahí lo del “efecto mariposa”. Recapitulemos: un sistema mínimamente no lineal que es muy sensible a su punto de partida.

Y ahora volvamos a la realidad. Por favor, tachen toda la discusión anterior y pónganse en los zapatos de quien ve solo una de estas series y, aun munido de datos de la historia de la serie, no conoce cuál es la regla que la generó ni la condición inicial. Si se le encarga que prediga los próximos valores, lo más posible es que el flamante analista “tire la toalla” y diga que estaba ante un comportamiento impredecible.

Entonces, las enormes dificultades predictivas que plantean los fenómenos regidos por estos sistemas caóticos tienen que ver con el “combo” dinámica, no linealidad e hipersensibilidad a condiciones iniciales.

La larga saga de la meteorología

Y si hablamos de complejo, caótico, azaroso y no lineal, la meteorología levanta bien alto la mano. Después de pasar un ratito analizando sistemas dinámicos que pueden tener comportamientos caóticos, es un buen momento para discutir "el" evento predictivo por excelencia, el que rige desde la aviación hasta el dilema de los domingos (¿prender o no prender el fuego para hacer un asado?): la predicción del tiempo.

Debe haber sido en octubre de 1993. Se venía mi primer invierno en el Midwest estadounidense, donde hice mi doctorado y trabajé algunos años, y el frío es cruel, criminal, mejor dicho: me ha tocado salir a dar clase con -30 ºC (tengo testigos). Habíamos cenado con un par de amigos en Chicago, a una hora y media de donde vivíamos, y se nos hizo tarde para el regreso. La noche cerrada nos sorprendió en la mitad del camino con una tormenta de nieve épica. Pude capear el temporal (literalmente) porque, siguiendo un oportuno instinto de supervivencia, sintonicé una radio local que emitía indicaciones sobre cómo manejar en esas condiciones extremas. Con miedo y prudencia llegamos a casa, exhaustos. Al día siguiente relaté mi peripecia a algunos compañeros, que con una sonrisa burlona me dijeron: "Eso te pasa por ser sapo de otro pozo, bienvenido al invierno del Midwest". Y me explicaron que, llegado el otoño tardío, nadie en su sano juicio encara la ruta sin chequear antes el pronóstico del tiempo, por la radio o la tele, en épocas de incipiente internet.

Ya vimos que un viejo chiste dice que la economía se inventó para que los meteorólogos parezcan precisos; todos recuerdan un asado frustrado por una lluvia que nadie previó, o haber cargado campera y paraguas al divino botón, por un pronóstico que jamás se cumplió. Al respecto, caben dos apreciaciones. Primero, no hay ninguna evidencia de que quienes se dedican a la meteorología sean vagos o mentes inferiores. Todo lo contrario, la carrera de Ciencias

de la Atmósfera (o similares, donde se estudia meteorología) requiere una sólida formación en física, matemática, computación, estadística, además de las especificidades de la materia. No es una carrera "de cómo darle la mano a un japonés" (como jocosamente decía un profe mío para referirse a la frivolidad imperante en otras disciplinas), sino que demanda conocimientos de frontera en lo más profundo de las ciencias naturales y exactas. No es por ahí: las fallas predictivas de la meteorología surgen de la brutal complejidad de las leyes que rigen el clima y el tiempo, no de las limitaciones de los meteorólogos.

Segundo, ¿en verdad andan tan mal los pronósticos meteorológicos? Creer o reventar, la meteorología de corto plazo funciona muy bien y ha mejorado considerablemente en los últimos tiempos. Veamos, ¿a quién le importa el tiempo en el corto plazo? Bueno, ¡a quién no! Son estos pronósticos los que conciernen tanto a la aviación como a que la gente sepa cómo vestirse, a la agronomía y al camionero que debe elegir una ruta apropiada para manejar seguro, a hacer un asado o no, a evitar las consecuencias de una inundación, y a que no te atrape una tormenta de nieve, como a mí en Illinois.

Tengo la impresión de que con los meteorólogos pasa algo parecido a lo que ocurre con los mozos de un restaurante o los referís de fútbol: la gente recuerda asimétricamente sus performances. Nunca escuché cantitos de cancha en favor del referí después de un partido en el que tuvo una actuación adecuada. Por el contrario, una mano no cobrada o un desliz en un offside que afecta a nuestro equipo gira todas las discusiones posteriores en torno al árbitro. Con los meteorólogos pasa lo mismo. Insultamos en arameo si un chaparrón imprevisto nos agarra sin paraguas, pero no tuiteamos emojis cariñosos cada vez que el pronóstico se cumple.

De hecho, el 17 de enero de 2013, la Cámara Argentina de Lavaderos Automáticos y Manuales de Automotores (Calama: sospecho que primero inventaron la sigla y después buscaron las palabras que la conforman) envió una furibunda carta de lectores al diario La Nación, que decía que los yerros en los pronósticos del Servicio Meteorológico Nacional (SMN) afectaban negativamente a su negocio. Más lejos fue Vladímir Puchkov, ministro ruso de Defensa Civil, Situaciones de Emergencia y Socorro en Casos de

Desastre, quien, en 2017, una semana después de que 16 personas falleciesen en Moscú por un temporal, sugirió aplicar multas a los responsables de los servicios meteorológicos cuyos pronósticos resultaran incorrectos o no fueran comunicados a tiempo. La iniciativa de Puchkov no prosperó. Los economistas silban bajito para no levantar la perdiz.

En fin, la creencia popular es que el pronóstico meteorológico no funciona. Hay formas científicas de chequear esta aseveración y las trataremos más adelante, no se impacienten. Mientras tanto, déjenme contarles una manera más casera y no menos efectiva, que me encantaría que implementaran. En los cursos de estadística para principiantes que dicto suelo invitar a los alumnos a que hagan un ejercicio similar: que elijan una "estadística" y la sigan o persigan por un tiempo, que indaguen cómo se construye, si funciona, si no, etc. Puede ser deportiva, política, financiera, meteorológica o de cualquier otra laya.

Daniela Parada, a quien ya mencionamos, tomó en serio esta sugerencia. Su pequeña hija Mila y ella misma se encargaron de ver si en verdad el pronóstico meteorológico pifia tanto. Lo que hicieron Daniela y su hija es exactamente lo que les pido a mis alumnos. Todos los días, durante seis meses, registraron en una planilla Excel el pronóstico de lluvia y, al día siguiente, si había llovido o no.

Antes de analizar los datos recabados por Daniela y su hija, déjenme contarles qué es lo que opina "lajente" (como está de moda decir en las redes sociales) al respecto. Hice una pequeña encuesta en Twitter, en la que pregunté: "Si 0% es que nunca acierta y 100% es que siempre lo hace, ¿qué opinás de la tasa de acierto del pronóstico del tiempo del SMN?". Escaloné las respuestas en cuatro intervalos: de 0 al 25%, del 26 al 50%, del 51 al 75% y del 76 al 100%. Acá les pido, por favor, que piensen tres cosas. Primero, en el estilo del programa Cien argentinos dicen, cuál creen que fue la respuesta a la encuesta. Segundo, cuál es su respuesta, y tercero, cuál les parece que fue el resultado de los datos de Daniela y Mila.

Empecemos por Twitter. Allí, la mayoría opina que la tasa de acierto del Servicio Meteorológico Nacional está en el intervalo 51-75%, o sea, que anda, pero más o menos.

Los resultados de Daniela dicen que la tasa de acierto es (maestro, redoble y platillazo, por favor): ¡un 97%! Sí, creer o reventar, casi siempre le pega al anunciar si en un día llueve o no. Por supuesto, esta encuesta casera es criticable y mejorable, hay varias cuestiones finas por resolver. En primer lugar, el pronóstico del tiempo no dice "llueve" o "no llueve", sino que habla de probabilidades de lluvias, de chaparrones aislados o de tormentas. En el capítulo 1 nos referíamos a la tensión que hay entre el hecho de que un fenómeno ocurra y sus chances de que eso suceda. Otras sutilezas se relacionan con preguntas en apariencia inocentes como esta secuencia: "¿Qué quiere decir que llovió? ¿Significa que cayeron más de tantos milímetros? ¿Que llovió en muchos lugares? ¿Se puede decir 'llovió en Buenos Aires' si solo cayeron unas pocas gotas en una plaza de Belgrano y no en el resto de la ciudad?". Otra cuestión de aparente relevancia apunta al horizonte del pronóstico: una cosa es decir que en el próximo milisegundo va a llover o no, y otra que va a llover en algún momento en las próximas veinticuatro horas. En definitiva, la encuesta es mejorable (y llevarla adelante es un muy interesante ejercicio), pero poco afecta a la contundencia de los resultados: en el corto plazo (un día, en este caso) el pronóstico del SMN funciona muy bien. Demasiado bien.

Complejidad y sistemas dinámicos en la predicción del tiempo

Ahora bien, ¿por qué es tan difícil predecir el tiempo? Por el carácter mismo de las leyes que gobiernan al clima y a los fenómenos meteorológicos, que tienen altas dosis de la complejidad y no linealidad de las que hablamos antes. No, no los voy a aturdir con la física de las nubes, los vientos y la lluvia, fundamentalmente porque los detalles se me escapan, como a la mayoría de los seres humanos.

En una jugosa y detallada monografía reciente que revisa lo acontecido con los pronósticos del tiempo en los últimos cien años, Stanley Benjamin y sus coautores dicen: “Todos los pronósticos, y especialmente los pronósticos de NWP, son problemas de valores iniciales”. La sigla NWP refiere a numerical weather prediction o, aproximadamente, predicción numérica del tiempo, tal vez el método más importante usado en pronósticos meteorológicos y “culpable” de las notables mejoras predictivas.

Como ocurrió con muchas cuestiones de la ciencia y la tecnología –y por las peores razones–, la Segunda Guerra Mundial fue un motor importantísimo para el desarrollo de la meteorología. Señala Benjamin en su ensayo que hasta 1939 los pronósticos del tiempo se basaban en inter- y extrapolaciones de mapas de confección manual, con un alto componente de subjetividad. La guerra y sus urgencias lo cambiaron todo. La aviación, las radiosondas, los radares y la irrupción de las computadoras posibilitaron que las complejas leyes de la física se aproximasen con una sabia mezcla de procesos numéricos “maridados” con datos. Y así toma forma el enorme edificio de la NWP: un conjunto de métodos físicos y matemáticos que, alimentados con datos de la realidad y gracias a la asistencia de computadoras, permiten simular fenómenos atmosféricos y predecirlos.

Es el avance arrollador en todas las dimensiones involucradas en estos modelos lo que está detrás de las mejoras en la confiabilidad

de los pronósticos de corto plazo. En relación con la capacidad predictiva de la meteorología, la versión científica de la cuentita que sacamos con los datos caseros recopilados por Daniela Parada está en la monografía de Benjamin. El “score S1 estandarizado” es una métrica estándar para cuantificar la capacidad predictiva de los modelos de NWP. En una escala que va del 0% (poco confiable) al 100% (predicción perfecta), este índice estaba en 20% en 1954, y fue creciendo de modo persistente hasta alcanzar valores de alrededor del 85% en 2018, en referencia a la capacidad predictiva de estos modelos para un horizonte de 36 a 72 horas. O sea:

las predicciones meteorológicas de corto plazo son muy confiables,

han mejorado de manera sustancial en los últimos cincuenta años.

En esta parte del relato es interesante mencionar que la meteoróloga argentina Eugenia Kalnay es autora de “la” referencia obligada sobre predictibilidad en cuestiones atmosféricas (citada recurrentemente por Benjamin), y protagonista de una historia trágica e inspiradora, porque sus enormes logros científicos ocurrieron después de haber tenido que abandonar el país en 1966, a continuación de la Noche de los Bastones Largos, e instalarse en los Estados Unidos. Y si de resaltar logros meteorológicos vernáculos se trata, es imposible no mencionar el caso de Celeste Saulo, que por su larga experiencia (incluido su paso por la conducción del SMN) en enero de 2024 fue elegida secretaria general de la Organización Meteorológica Mundial (OMM), el organismo autorizado de Naciones Unidas que estudia las cuestiones del tiempo, el clima y el agua.

Las mejoras contundentes en la confiabilidad de los pronósticos del tiempo están relacionadas con los avances físico-matemático-computacionales involucrados en la construcción de modelos operativos de los fenómenos atmosféricos, y también con la

disponibilidad de datos precisos que permiten alimentarlos. “Alimentarlos” tiene que ver con fijar con precisión la posición inicial de estos modelos (similares a las “condiciones iniciales” de nuestra discusión anterior), piedra fundamental para realizar simulaciones y pronósticos. A este proceso se lo conoce como asimilación de datos, y es el portal que une la realidad de la atmósfera con la maraña de ecuaciones y computadoras que, usando estos datos como insumos, producen simulaciones y pronósticos.

Para hablar al respecto, les presento a otra destacada meteoróloga argentina, María Eugenia Dillon, del SMN, que es toda una experta en los detalles de esta etapa crucial del proceso de pronóstico. En 2016, María Eugenia fue reconocida por la OMM con el Premio a la Investigación para Jóvenes Científicos por sus trabajos en procesos de asimilación de datos.

Habrán notado que intento usar muy puntillosamente el término “tiempo” y no “clima” para referirme a la temática de esta sección. Cuando entrevisté a María Eugenia para este libro, fue lo primero que me corrigió: “Clima no; los pronósticos son del tiempo, no del clima, te lo dicen en la facultad el primer día”. Y sí, la meteorología tiene sus especificidades, como ahorro e inversión en economía, propaganda y publicidad en marketing, o poder y autoridad en ciencia política. Acerca de su formación, me dijo: “Soy 100% atmósfera”, que en su versión no abreviada quiere decir que es licenciada en Ciencias de la Atmósfera, recibida en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, donde también completó su doctorado. Como les anticipaba, en la actualidad, María Eugenia se desempeña en el SMN, institución que vela por las relaciones de la comunidad con el tiempo y el clima (¡lo aprendí!), que incluye la actividad de pronóstico meteorológico y la observación continua de la atmósfera usando conocimiento científico de frontera.

Me cuenta María Eugenia que ella misma tiene dificultades para explicar a sus familiares y amigos qué quiere decir “un 90% de probabilidades de chaparrones aislados”, cuestión no menor que nos ha demandado bastante energía en el capítulo 1 y que, en el contexto de los modelos de NWP, significa que llovió en el 90% de

las simulaciones computadas con los poderosos modelos de la meteorología.

Yo era el rey de este lugar: IA para la predicción del tiempo

Y aun cuando la performance de los modelos de NWP tuvo muy considerables mejoras, es natural indagar qué tiene para aportar al respecto la IA. En las casi setenta páginas de la milimétricamente precisa monografía de Benjamin y sus coautores, publicada en 2019 y en la que revisan cien años de progresos en la predicción del tiempo, solo al final y de forma muy tangencial aparece una referencia a la IA: “Es evidente que la inteligencia artificial (esto es, deep learning) y las técnicas de machine learning serán usadas rutinariamente para integrar pronósticos en los procesos decisorios”.

Vale cuestionarse aquí si las mejoras predictivas en la meteorología se trasladan con igual confiabilidad a las predicciones sobre el futuro de la predicción de la meteorología. Un reciente estudio de Remi Lam y un equipo de Google, publicado en la prestigiosísima Science, muestra que, como cantaba el Indio Solari, en meteorología parece que el futuro efectivamente llegó hace rato.

En la jerga de la estadística predictiva, NWP es un modelo estructural, un conjunto de ecuaciones que representan el funcionamiento de un sistema, en este caso, de la atmósfera. Esta estructura es provista por la física. Dicho de otra forma, un modelo estructural surge de entender las causas, los mecanismos y las consecuencias que operan detrás de la meteorología, y expresarlas en ecuaciones matemáticas, que luego son puestas a interactuar con datos de la realidad para realizar simulaciones, evaluar el impacto de alterar el sistema y, a los fines de la temática de este libro, realizar predicciones. Desde este punto de vista, el modelo NWP parte de entender (parcialmente o no) cómo funciona el mundo, lo representa matemática y computacionalmente a través de un modelo que se usa para diversos propósitos, uno de ellos es predecir.

La IA procede de otra forma: aun a riesgo de sobresimplificar esta

delicada cuestión, anotemos que estudia copiosas cantidades de datos con métodos hiperflexibles, que serán tema del capítulo siguiente, para encontrar patrones en los datos que permitan realizar pronósticos. A diferencia de los modelos estructurales, los modelos de la IA no parten de entender el proceso subyacente, sino de buscar regularidades en los datos. Veamos esta suerte de ejemplo brutal: sobre la base de datos de paraguas y lluvias, la IA buscaría una forma de pronosticar lluvia con datos de uso de paraguas, sin proponerse explicar por qué llueve ni cuál es el rol que cumplen los paraguas en hacer llover, aun sabiendo que la respuesta es "ninguno". Hablaremos del nexo entre explicar y predecir en el capítulo 10.

La IA la tenía difícil. Los éxitos del NWP habían levantado la vara bien alta en relación con la capacidad predictiva de corto plazo de la meteorología, y por eso causó tanto revuelo el reciente paper de Remi Lam y su equipoo de Google. En ese estudio informan las experiencias de un modelo que llaman MLWP, por machine learning weather prediction (predicción del tiempo basada en aprendizaje automático). A diferencia de NWP, MLWP no se basa en las ecuaciones dinámicas de la física de la atmósfera (o, al menos, no directamente), sino en la brutal cantidad de datos históricos disponibles sobre el tiempo y en una arquitectura de redes neuronales (y otra vez me toca remitirlos al capítulo 7). En este caso específico, se trata de una "red neuronal gráfica" a la que, calculo, ya extenuados por haber puesto toda la energía creativa en las cuestiones computacionales del método, bautizaron como "GraphCast". Si una de las más grandes cadenas internacionales de comida rápida se llama (en inglés) El Rey de las Hamburguesas, están perdonados (El Rey de los Pronósticos no desentonaba demasiado).

El sistema se alimenta del estado del tiempo en un momento, con una resolución de unos 28 km^2 a la altura del Ecuador. Y no solo se alimenta: cual monstruo de película de terror futurista, GraphCast se "retroalimenta" de sus propios pronósticos, lo que le permite realizar predicciones a más largo plazo. El modelo fue entrenado con datos históricos de un lapso de veintinueve años (1979-2017) del Centro Europeo para Pronósticos del Tiempo de Rango Medio (Ecmwf es su sigla en inglés). La evaluación predictiva se hizo con

datos de 2018 en adelante, que le fueron ocultados al modelo para ver si podía predecirlos correctamente, una técnica habitual para evaluar algoritmos predictivos.

Los resultados son espeluznantes. GraphCast tiene mejor capacidad predictiva que el principal competidor usado por el Ecmwf, basado en los modelos de NWP. Concretamente, en la tarea de pronosticar 1380 variables meteorológicas, GraphCast le gana al método estándar en el 93% de los casos. Además, GraphCast puede operar a una resolución más fina que los métodos de NWP y a una frecuencia más alta –cada una hora– en comparación con los intervalos de seis horas en los cuales está disponible la predicción del Ecmwf. Con una mezcla de orgullo y vehemencia, Remi y sus coautores dicen: "Creemos que esto marca un punto de inflexión en la predicción meteorológica", si bien son cautos en cuanto al futuro cercano de los métodos, al sostener que

nuestros resultados no deben ser interpretados como un reemplazo de los métodos tradicionales de pronóstico, que han sido desarrollados por décadas y testeados rigurosamente en muchos contextos del mundo real, […] sino que deben ser interpretados como evidencia de que los métodos de MLWP se encuentran ya en condiciones de alcanzar los desafíos de los pronósticos del tiempo y tienen el potencial de complementar y quizás mejorar los mejores métodos actuales.

En síntesis, en meteorología coexisten dos circunstancias:

todavía es virtualmente imposible predecir si va a llover un 30 de septiembre de acá a dos años,

muchos fenómenos meteorológicos de corto plazo no solo son

predecibles, sino que la confiabilidad de estos pronósticos ha aumentado de manera considerable.

Queda abierta la pregunta de cuánto puede aportar la IA para mejorar estas cuestiones, en el corto y en el largo plazo.

¡Ghana gana el Mundial! Algoritmos y datos para predecir goles, gambetas y medallas

Y tras haber invertido tiempo en revisar cómo funciona este asunto de las simulaciones, estamos listos para meternos en la trastienda de un campo en el que los pronósticos valen oro, tanto por cuestiones emotivas como monetarias: el de los deportes.

Cada cuatro años el mundo se detiene expectante ante la justa deportiva por excelencia: el Mundial de Fútbol. Una característica crucial del "juego bello" es que tiene una dosis exacta de azar. Más, sería como ver a alguien tirando un dado, y menos, como ver a alguien haciendo una multiplicación. Es esa certera mezcla de fortuna y talento lo que está detrás de su popularidad, que atraviesa los continentes, las clases sociales y el paso del tiempo.

Para analizar cómo se elaboran los pronósticos del Mundial de Fútbol, déjenme proponerles una versión distópica (y espantosa) de este evento, en la que el ganador de un partido se dirime con una app de un celular. En este "Mundial Mad Max", son las 11.59 del domingo 18 de diciembre de 2022. El corazón de los argentinos parece querer salirse de la caja torácica. El árbitro polaco Szymon Marciniak, profesional y adusto como cualquiera en ese rol y con ese apellido, está igual de nervioso, pero disimula bien. Exactamente a las 12.00 hace sonar su silbato y, siguiendo a rajatabla el protocolo de la FIFA, y de acuerdo con el sorteo previo, Lionel Messi, favorecido por el tradicional lanzamiento de moneda, procede a presionar el botón de la app de marras, que da como ganador del partido (y del Mundial) a… ¡la Argentina! La parcialidad rioplatense enloquece, se suceden escenas de una película épica, la alegría invade las calles de Ushuaia a La Quiaca, Francia, seg…, seguro que sufre, como mi editora, que lee esta frase y cree que no evitaré el más trillado de los chistes mundialistas.

A fin de entender qué hacen los algoritmos que predicen el Mundial, supongamos que se sabe fehacientemente que, en los dos

primeros partidos, las chances que tiene Argentina de vencer a sus rivales son del 0,5 para el primero y del 0,7 para el segundo partido. ¿Cuáles son las chances de que Argentina gane los dos partidos? De aplicar toscamente las reglas de la probabilidad, las chances son del 0,35, cifra que surge de multiplicar 0,5 por 0,7. Como vimos en el capítulo 1, este cálculo presupone partidos independientes, es decir que, por ejemplo, el resultado del primer partido no influye en el segundo. Nada menos cierto en el fútbol; todos hemos visto cómo la moral de un equipo se altera luego de un resultado inesperado, que afecta al siguiente partido. Pero, para facilitar el análisis, supongamos que los partidos son independientes; ya lidiaremos con este problema. Entonces, las chances de que Argentina gane sus dos primeros partidos son del 0,35. No hay que hacer ninguna simulación, solo buscar la fórmula en un libro de probabilidad, o preguntársela a Roccatagliata, que todo lo sabe.

Procedamos ahora por simulación. He aquí un algoritmo. Pidámosle a la computadora que tire una "moneda computacional", que sale cara con chances del 0,5. Si sale cara, Argentina "gana"; esto simula el primer partido. Ahora le pedimos que tire otra moneda, pero cargada para que salga cara con chances del 0,7, y si sale cara, Argentina gana de vuelta. Entonces, para sacar la cuenta final, en cada paso le pedimos al algoritmo que tire estas dos monedas, y si sale dos veces cara, Argentina gana los dos partidos. Luego le pedimos que repita el paso anterior miles, millones de veces, y pasamos a contar la proporción de veces que salió cara dos veces: ahí está la estimación de la chance de que Argentina gane dos partidos. Yo lo hice con mi computadora 1000 veces (bueh, lo hizo Roccatagliata, que para eso está), y Argentina ganó 348 veces; ergo, la probabilidad estimada de que ganara los dos partidos fue del 0,348. Muy parecido a 0,35, ¿no? Sí, como con lo del arroz y el círculo en el capítulo anterior.

OK, ahora saltemos al Mundial de verdad. No hay tres equipos, sino treinta y dos; además, como sabemos, la estructura del campeonato no es simple. Hay una fase de grupos, en la que todos juegan contra todos; luego vienen los octavos, los cuartos, la semifinal y la final. Solo en la primera fase es posible saber de antemano quién juega con quién. A partir de ahí, los partidos dependen de quién clasificó

primero o segundo y quién pasó de ronda. O sea, en el Mundial de verdad es casi imposible apelar a una formulita sencilla que mida cuáles son las chances de que Argentina (o Francia, o México, o cualquier equipo) gane el torneo.

¿Cómo se hace en la práctica? Sí, adivinaron, se repite la misma estrategia de vago a la que apelamos antes para ver si Argentina ganaba dos partidos, pero con todo el Mundial. O sea, en el algoritmo, el ganador de cada partido surge de tirar una "moneda computacional" cargada para que refleje las chances de que un equipo le gane a su contrincante. Esto se hace por separado para cada zona, donde aparecen los que clasifican, tras lo cual se vuelven a jugar partidos computacionales y así sucesivamente, hasta que se llega a la final. Así, en estos "mundiales computacionales", en una de las corridas que son millones, puede que Ghana derrote a Portugal, luego a Corea del Sur y también a Uruguay, pase a octavos, luego a cuartos, venza en la semifinal y luego liquide a Francia en la final, y Ghana gane el mundial computacional. (Y menos mal, porque ya se me habían agotado los sinónimos de "gana"). Festeja la parcialidad ghanesa, Francia, seg… –perdón– seguirá soñando con tener más suerte en otra iteración.

Con las salvedades del caso, esto es lo que hace un algoritmo profesional de pronósticos de mundiales. La computadora corre el Mundial millones de veces y calcula en qué proporción de estos mundiales ficticios ganó tal o cual equipo. En el de Qatar, este método fue implementado por un equipo del Instituto de Cálculo de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, equipo liderado por Guillermo "Willy" Durán, un matemático de prestigio y renombre internacionales, fanático del fútbol y los deportes.

El proyecto, denominado "301060" en homenaje a Diego Armando Maradona (Roccatagliata sabe qué significa el número, seguro que ustedes también; si no, bien pueden buscarlo en Google), es una auténtica usina de simulación de partidos de fútbol. Por supuesto, los partidos se dirimen sobre la base de procedimientos más sofisticados que tirar una "moneda computacional", pero, en espíritu, la lógica es la descripta antes, en el sentido de simular los resultados sobre la base de chances. Para mostrar la relevancia de

simular, el sitio web de 301060 dice que la ruta analítica requeriría lidiar con más de mil millones de combinaciones de resultados en tan solo un grupo de la primera fase, situación en la que las simulaciones son la única alternativa práctica. El sitio del proyecto contiene una descripción muy pormenorizada de todos los mecanismos, datos y supuestos usados para simular el Mundial, en un lenguaje accesible al público y también convincente para los expertos.

Por ejemplo, la página web de 301060 –que se anticipó a "preguntas posibles"– presenta esta situación: "¿Cómo se calcula la probabilidad de que […] Argentina le gane 2-0 a México?". En su respuesta, dice:

Una vez que tenemos la base de datos con los partidos, […] mediante un modelo matemático con ayuda de la computación, calculamos la "fuerza" de ataque y de defensa de ambos equipos. De esta forma, podemos obtener las probabilidades de que Argentina le haga 2 goles a México y de que no reciba ninguno. Técnicamente la probabilidad de que Argentina le haga x goles a México sigue una distribución de Poisson con parámetro la fuerza de ataque de Argentina multiplicada por la fuerza de defensa de México (cuanto más alto es este valor, más débil es la defensa, a diferencia de lo que pasa con la fuerza de ataque), mientras que la probabilidad de que México le haga y goles a Argentina sigue una distribución de Poisson con parámetro la fuerza de ataque de México multiplicada por la fuerza de defensa de Argentina. Se calculan todos los posibles resultados de ese partido (acotando razonablemente el número de goles que un equipo puede hacer o recibir).

Seguro que detalles como los referidos a la distribución de Poisson se les pasaron, pero la explicación es clara en que hay una forma precisa (que los expertos entienden) de computar estas chances y luego simularlas con una computadora.

La simulación se lleva a cabo con un millón de replicaciones y, antes del Mundial, el 301060 le asignaba a Brasil un 26,74% de

chances de ganar. Seguía Argentina con un 11,08% y seg…, perdón, tercero Francia, con un 10,70%.

Más allá de la relevancia y la utilidad de 301060, se trata de un claro ejemplo de buena ciencia aplicada. Lejos del pulpo Paul o los pálpitos de los seudoexpertos, el proceso detrás de 301060 es transparente; cualquiera puede replicarlo (con la pericia matemática y los enormes recursos computacionales que requiere, por supuesto), discutir sus supuestos y mecanismos y, por qué no, proponer alternativas y mejoras.

Así avanza la ciencia. Y así también avanza el fútbol profesional, cuya quirúrgica mezcla de pericia y azar logra detener el tiempo cada cuatro años y hacer latir fuerte los corazones de argentinos, de italianos, de ghaneses y… de franceses también, seg… según me consta.

Volver del futuro

Los pronosticadores científicos no se juegan por sus pronósticos, sino por la forma en la que se producen, que es clara (por lo menos, para los expertos), reproducible y consciente de sus alcances y limitaciones. En este capítulo discutimos estas ideas:

Detrás de las notables mejoras en los pronósticos meteorológicos de corto plazo está el enorme edificio de la física, las certezas de la matemática y la potencia de las computadoras y los datos. Detrás de los fracasos, la enorme complejidad del clima.

Lo mismo ocurre con el fútbol y el deporte competitivo, cuya aparente impredecibilidad tiene que ver con que en sus mismísimas entrañas se esconde una sabia mezcla de lógica predecible y aleatoriedad, esa que hace que, si bien es altamente probable que un seleccionado profesional le gane a un equipo primerizo en los mundiales, nada está dicho hasta que el árbitro no dé la pitada final de su silbato, como ocurrió en el partido de Argentina versus Arabia en Qatar.

Si los capítulos iniciales se focalizaron en el rol del azar, las historias de este capítulo suman una capa adicional a esta ensalada de impredecibilidad: la proveniente de la complejidad, entendida como resultado de interacciones no lineales y dinámicas.

7. Armas de predicción masiva: los pronósticos y su caja de herramientas técnicas

Entre las preguntas más molestas que le pueden hacer a alguien que se dedica a la ciencia de datos está "¿qué opinás de XXX?", en la que XXX no es ninguna película "para adultos", sino algún método predictivo, como Lasso, boosting o bagging, que más que a técnicas suenan a tipos de gimnasia, como zumba o spinning, con un instructor que grita todo el tiempo. En fin.

En primerísimo lugar comento que juzgar una técnica sin especificar para qué se la demanda es un tanto inconducente. Las opiniones acerca de un martillo difieren según se trate de clavar clavos, de quitarlos, de atornillar tornillos o de ablandar milanesas. Ya mencionamos la lograda frase de George Box ("todos los modelos están mal, pero algunos son útiles"). Por eso, si vamos a referirnos a los métodos predictivos, primero deberíamos explicitar qué pretendemos hacer y cómo medir si lo estamos haciendo bien, y luego explorar y juzgar las técnicas. Y ese es el objetivo y el orden que seguiremos en este capítulo aventurero, que es una visita guiada a la "ferretería de las predicciones".

Así que consigan lápiz y papel (o una tablet, para sonar modernos) que ya parte el tren de la alegría de las herramientas predictivas. Presten mucha atención a la elección de las técnicas, porque "cuando tenés un martillo, todo parece un clavo", como decía Abraham Maslow (sospecho que mientras martillaba tornillos).

El campeón de los pronósticos

Por escándalo, el promedio es la estadística más popular de todas, es el "Rasguña las piedras" de la profesión, "una que sabemos todos". Todo el mundo sabe cómo se calcula, nadie googlea qué es un promedio y la enorme mayoría de los guarismos de uso cotidiano son un promedio, aun cuando no lo parezcan.

A ver. El promedio de notas ocupa un lugar central en la educación, todos hemos lidiado con el promedio de las notas de una materia. Pero la mayoría de las cifras cotidianas también son promedios. Por ejemplo, la tasa de pobreza, y, si vamos al caso, cualquier proporción. Cuando por la tele dicen que la tasa de pobreza es del 40%, significa que, de todos los hogares encuestados, el 40% es pobre. OK, pero… ¿y el promedio dónde está? Fácil. Supongamos que en la encuesta había 3000 hogares. Si la tasa de pobreza es del 40%, es porque 1200 hogares eran pobres. Doy más detalles. Si el encuestador anota en una planilla un 1 cuando el hogar es pobre y un 0 cuando no, y sumamos todos los ceros y los unos, veremos que la suma da, sí, 1200. Si dividimos 1200 por 3000, da 0,4, o sea, el 40%: la tasa de pobreza es también un promedio, de los ceros y unos que usamos para denotar pobres y no pobres. En esta línea, el porcentaje de votantes a tal o cual candidato también es un promedio.

En un libro sobre pronósticos, como este, el promedio (y sus variantes) ocupa un lugar central, por un resultado importantísimo de la estadística: en algunas condiciones, el promedio es la mejor predicción que se puede hacer.

Volvamos a un ejemplo del capítulo 1. Supongamos que en una clase hay 40 alumnos, con una edad promedio de 15 años. Ahora supongamos que los alumnos todavía no han llegado y que, de aburridos nomás, jugamos a predecir la edad del primer alumno que entra al aula. ¿Cuál es la mejor predicción? Bueno… 15 años, el promedio.

No es difícil demostrar matemáticamente que el promedio es el valor que mejor representa a una colección de números, el que está “más al medio” de todos. La clave está en presuponer que la noción de “lejos” entre un número y otro se mide tomando el cuadrado de la diferencia entre ellos (como vimos, eso también contempla a los números negativos).

Regresión: el automóvil de la ciencia de datos

Argumentamos en el capítulo 4 que las probabilidades condicionales son mejores para predecir que las comunes. Y resulta que este razonamiento se traslada a los promedios. Es decir, los "promedios condicionales" predicen potencialmente mejor que los promedios simples.

Por ejemplo, si en el populoso barrio porteño de Belgrano el precio promedio de un departamento es de, digamos, US$120.000; entonces, a falta de más información, la mejor predicción que se puede hacer de cuánto cuesta cualquier departamento en Belgrano es US$120.000. Si a uno le interesase un departamento pequeño, el precio predicho debería ser algo menor, como en el caso de la librería Daniel del capítulo 4.

En nuestra jerga, US$120.000 es el promedio no condicional de los precios de los departamentos en Belgrano. Pero si nuestro interés recae en un departamento pequeño en Belgrano, el mejor predictor no es el precio promedio, sino "el precio promedio de los departamentos pequeños de Belgrano". A esta segunda versión se la llama promedio condicional y, forzando viejos argumentos, se puede demostrar que este será el mejor predictor de los precios de los departamentos pequeños de Belgrano.

Supongamos que ahora nos interesa predecir el precio de un departamento de 50 m^2 de superficie. Entonces, el mejor predictor ya no es el promedio de todos los departamentos, sino el promedio de los departamentos de 50 m^2. Si seguimos esta lógica, podríamos afinar la puntería aún más en caso de que fuésemos muy estrictos con nuestras preferencias. Por ejemplo, si nos interesara un departamento de 50 m^2, con una habitación, vista al frente, calefacción central y en un piso por arriba del tercero, calculo que ya se dieron cuenta de cuál es el mejor predictor del precio de ese departamento: el precio promedio de todos los departamentos que satisfacen estrictamente esas características.

No es difícil demostrar que los promedios condicionales predicen mejor que los no condicionales. Entonces, en la medida de lo posible, a los fines predictivos es mejor usar promedios condicionales que no condicionales. A la técnica estadística que se usa para calcular promedios condicionales se la llama regresión. Sí, un nombre esotérico, que remite a psicoanálisis y cosa rara; de hecho, el término ya apareció en el capítulo 3 cuando hablamos de regresión a la media.

Stephen Stigler, el principal historiador de la estadística, dijo que el análisis de regresión es "el automóvil de la estadística", y opino que no exagera un centímetro. Déjenme explicarles por qué. Parece fácil calcular promedios condicionales. La receta es la siguiente:

hagan una lista de las características que les interesan ("características en las que condicionar", es la jerga utilizada);

de los datos originales, quédense con los que satisfacen esas características y

calculen el promedio de la variable que les interesa para los datos del paso

En el ejemplo de los departamentos, si nos interesan los de 50 m^2, con una habitación, vista al frente, calefacción central y en un piso por arriba del tercero, de los datos disponibles, aislamos los que tienen estas características y luego calculamos el precio promedio de esos departamentos. No parece muy difícil.

Pero hay un problema con esta receta: si uno se pone muy puntilloso con los "condicionales", posiblemente no haya datos en la muestra original con los cuales calcular el promedio. Y cuanto más detallista sea uno con las características, menos datos debería

haber (si alguno) para calcular el promedio condicional.

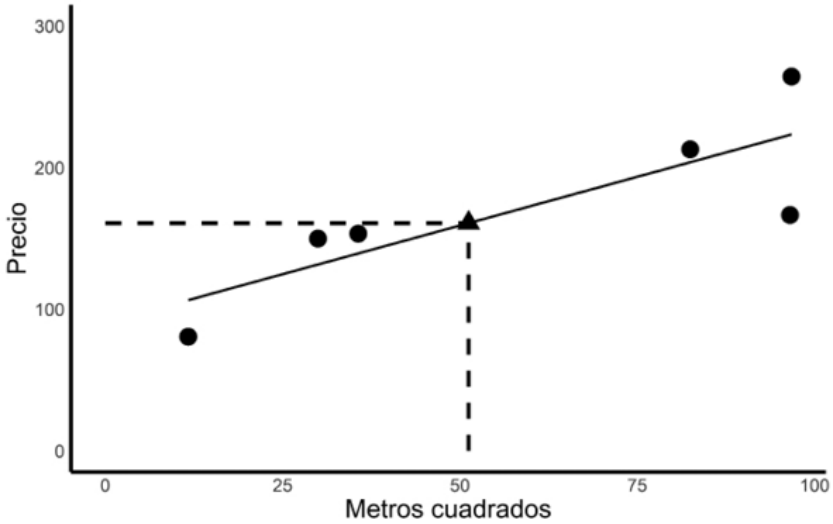
Por ejemplo, si un comprador meticuloso (al que llamaremos Jorge) quiere predecir el precio de un departamento de 51,23 m^2, es bastante posible que no haya ningún departamento con esa exacta medida, por lo cual no se puede seguir la receta descripta. Misma cosa si Jorge (un molesto, digámoslo) quiere predecir el precio de un departamento de 51,23 m^2, de tres ambientes, con cocina triangular, vista al frente, en un piso impar y en una calle cuyo nombre empieza con la letra J.

Vayamos al caso del departamento de 51,23 m^2. Es muy posible que no haya en los datos departamentos con esa misma superficie, pero quizás haya de 50 m^2 o de 55 m^2. Entonces, "tomando prestado", podríamos "tunear" el precio de uno de 50 m^2 para arriba un poquito, o el de uno de 55 m^2 para abajo, y así aproximar al de 51,23 m^2 que pretende el molesto de Jorge.

Bueno, esta negociación es exactamente lo que hace el análisis de regresión. ¿Y por qué tanta bambolla? ¿Qué da justificación a eso de "el automóvil de la estadística"? En realidad, con algunos supuestos, la regresión permite predecir los precios de cualquier departamento que al denso de Jorge se le ocurra, se observen en la práctica o no. Magia.

Les muestro en un dibujito lo que hace el análisis de regresión lineal, para un ejemplo de juguete con solo seis datos. En la figura 7.1, en el eje horizontal están los metros cuadrados de los departamentos, y en el vertical, los precios. La línea que pasa por entre los puntos es la recta de regresión. Esta recta modela razonablemente bien la relación entre tamaño y precios, fíjense que pasa cerca de todos los puntos. Y, además, predice algo coherente: departamentos más grandes salen más caros.

Figura 7.1



Ahora podríamos usar la recta para predecir el precio de cualquier departamento. Por ejemplo, el que le gustaría a Jorge es uno de 51,23 m^2. En los datos no hay ninguno con esa medida exactamente, pero si seguimos la recta que representa a todos los puntos, el modelo predice que ese departamento sale US$160.880, un poquito más que los de 50 y menos que los de 55 m^2. Nuestra ganancia con el análisis de regresión es que pudimos predecir el precio de un departamento que nunca vimos en los datos. Es decir, primero fijamos un tamaño (los ansiados 51,23 m^2, en el eje horizontal) y luego predijimos su precio con la recta (en el plano cartesiano, el valor en el eje vertical del "triángulo", US$160.880).

Entonces, el análisis de regresión elige la mejor recta que representa a todos los puntos. Si se da por sentado que esta recta representa bien tanto a los departamentos que están en la base de datos como a los que no, se puede usar para predecir el precio de cualquier departamento.

¿Dónde está la magia, lo que hace crucial al análisis de regresión? En que pudimos decirle a Jorge cuál sería el precio del departamento que le interesa, aunque en los datos no haya ninguno que tenga exactamente la superficie que él busca. Esta capacidad de "rellenar agujeros informativos" es una de las razones que explican

la enorme popularidad de los métodos de regresión.

Y la cosa no termina aquí. Resulta que, matemática mediante, es posible extender la idea de regresión a cuantas variables uno quiera y, así, predecir el precio de un departamento de 51,23 m^2, de tres ambientes, con cocina triangular, vista al frente, en un piso impar y en una calle cuyo nombre empieza con "J". Eso, siempre que dispongamos de una base de datos con la medida de los departamentos, la cantidad de ambientes, la forma de la cocina, su orientación, el piso y en qué calle está ubicado, cosa que cualquier inmobiliaria tiene. A la circunstancia en la que se intenta predecir una magnitud sobre la base no de una sino de varias variables se la llama regresión múltiple.

En definitiva, ¿cómo se calcula la recta de regresión o, en términos generales, la regresión múltiple? Bueno, ese es otro libro, o el objeto de estudio de cualquier curso de estadística básica. Si cursaron una materia de estadística y no les explicaron algo de regresión, sáquenle una foto a este párrafo y vayan a Defensa del Consumidor a que les reintegren la plata.

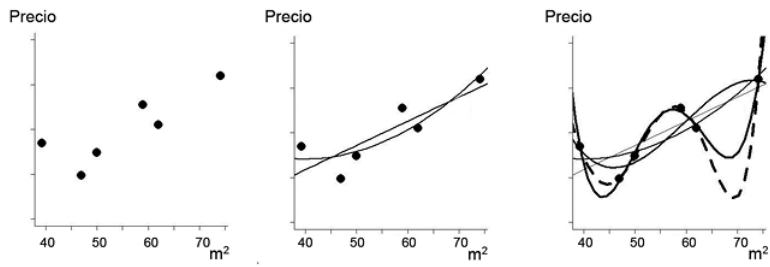
Ya nos quedó claro: el análisis de regresión es "la" máquina de sacar promedios condicionales, una forma eficiente de explotar la información disponible, la piedra angular de machine learning y la ferretería de la predicción. Stigler tiene razón. Y si Henry Ford fue quien puso el mundo sobre ruedas, Gauss y Galton (los padres fundamentales de la idea de regresión) son los Henry Ford del análisis de datos.

El que mucho abarca, poco aprieta: sobreajuste

En la sección anterior pasamos una recta por los datos de precios y tamaños de departamentos. ¿Y por qué una recta, eh? Bueno, bueno, si me lo reclaman así, pasemos ya mismo a una curva. Por ejemplo, la rama positiva de una parábola como las que vieron en el secundario. ¿Y por qué una parábola y no algo más complicado? OK, ¿quieren algo más complicado? Ahí vamos.

Fíjense en los siguientes dibujos, nuestra figura 7.2. Empezamos con la recta, seguimos con la parábola y después continuamos con curvas cada vez más flexibles (o bien: complicadas). Y fíjense muy bien lo que nos pasó en este viaje hacia la complejidad. Cuanto más complicada (flexible) es la curva, más cerca tiende a pasar de los datos. Así, la curva más complicada de todas pasa exactamente por todos los puntos. ¡Ganamos! Mientras celebramos, no demoran en aparecer las malas noticias. A partir de cierto momento, las curvas pasan muy cerca de los datos, pero fuera de ellos hacen cualquier disparate. Peor aún, predicen bien los datos que tenemos y horrible los que no.

Figura 7.2



Esta es una idea fundamental en machine learning: los modelos que funcionan muy bien dentro de los datos tienden a funcionar muy mal fuera de ellos. Y este es un problema gravísimo, porque nuestro propósito es predecir cosas que no vemos, como el departamento que quiere Jorge.

A este problema se lo llama de sobreajuste (por overfit, en inglés): en pos de pasar cada vez más cerca de los datos disponibles, una curva cada vez más flexible funciona mejor, pero al costo de empezar a predecir muy mal por fuera de los datos disponibles.

Resulta que la jerga que aprendimos en el capítulo 5 describe con precisión lo que está pasando: modelos muy complejos son poco sesgados (pasan cerca de los datos), pero tienen alta varianza (son muy erráticos). La tarea más importante de machine learning, y de las predicciones, pasa por "tunear" exactamente el balance entre sesgo y varianza, de modo de que, en conjunto, una herramienta prediga bien. Usando la jerga aprendida, predecir bien es minimizar el error cuadrático medio, la suma del sesgo (al cuadrado) y la varianza. Al hecho de que el sesgo guarda una relación inversa con la varianza se lo conoce como trade-off (compromiso) sesgo-varianza. ¿Quieren menos sesgo? Tienen que aumentar la complejidad, lo que aumenta la varianza. ¿Quieren menos varianza? Tienen que bajar la complejidad del modelo, lo cual aumenta el sesgo.

Este es el principal desafío de las versiones modernas del análisis de regresión. Con más datos (big data), uno se envalentona y considera modelos más complicados (machine learning, IA), pero el trade-off sesgo-varianza asoma su espantosa cabeza: si complicamos demasiado el modelo, el sesgo baja pero la varianza aumenta. Encontrar el "punto caramelo" de uno y otra es tarea a la cual destina buena parte de su energía el campo de machine learning-IA.

La profundidad del ChatGPT: redes neuronales y deep learning

Y si el análisis de regresión es el automóvil de la estadística moderna, deep learning es el DeLorean de Marty McFly de la IA. Pónganse casco, dejen un teléfono de referencia y ajústense los cinturones, que ahí vamos, a lo profundo.

Si regresión remite a psiquis e hipnotismo, deep learning evoca lo mismo, pero en versión Mad Max. Deep learning es el principal motor detrás de tecnologías disruptivas como ChatGPT y mucho de la IA moderna, y de un potencial enorme. Hay un punto de partida, que también contribuye a la mística: la llamada maldición de la dimensionalidad. ¡Flor de película de terror futurista podríamos hacer con toda esta terminología!

Vimos en la sección anterior que hay una complejidad óptima de los modelos de regresión. Los muy simples dan lugar a sesgos; los demasiado complicados, a mucha imprecisión. Ahora bien, esta lógica de "compromiso" entre sesgo y varianza es para una cantidad fija de datos. Por otro lado, dejando el modelo fijo, más datos es mejor.

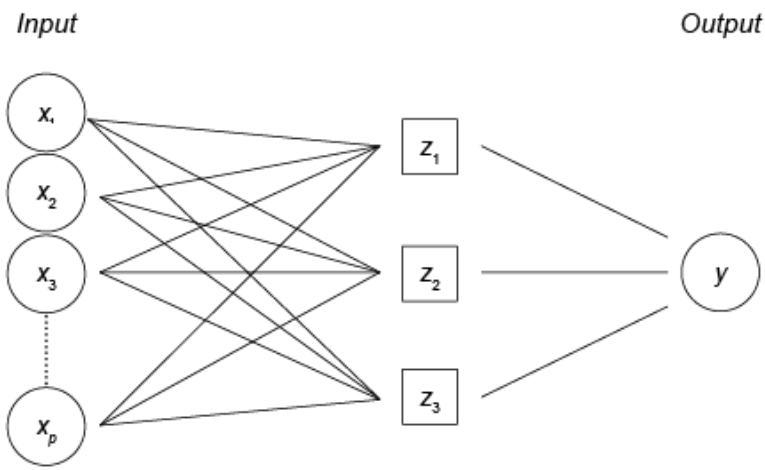
Resulta tentador pensar que uno podría considerar modelos más complejos a medida que crece la cantidad de datos disponibles. Lamentablemente, la maldición de la dimensionalidad dice, en términos muy generales, que si aumentamos la complejidad a cierta tasa, la cantidad de datos que debemos agregar para que la imprecisión no aumente tiene que crecer a una tasa muchísimo mayor. Por ejemplo, si el modelo se complica el doble, la cantidad de datos que hacen que la imprecisión no aumente tiene que aumentar mucho más que el doble.

Esto pone un freno al aumento desmedido de la complejidad de los modelos, que, sin los debidos cuidados, se da de bruces con la maldición de la dimensionalidad (que a esta altura parece una película de Freddy Krueger).

Desde esta perspectiva, deep learning es una suerte de modelo de regresión hiperflexible que, por su estructura, negocia adecuadamente la maldición de la dimensionalidad. Es capaz de aumentar mucho su complejidad sin que la limitación de los datos disponibles se devore sus promesas de mejora en la capacidad predictiva.

Deep learning es un derivado de las redes neuronales. En términos muy genéricos, una red neuronal es una suerte de "ejército" de modelos computacionales simples que, puestos a interactuar, son capaces de funcionar como un modelo hipercomplejo. La figura 7.3 nos ayudará a pensar esta cuestión de redes neuronales y deep learning.

Figura 7.3



Miren este dibujo, de derecha a izquierda. Una vez más, supongamos que el objetivo es predecir el precio de un departamento (y) sobre la base de un montón de características observables (tamaño, número de habitaciones y de baños, si tiene seguridad privada, etc.). Denotemos estas características como x1, x2, …, xp, (y aquí p es el número de características observables).

El gráfico muestra las características (x) como círculos en la primera columna, y el output (resultado), y, como el único círculo, a la derecha. En el modelo de regresión analizado antes, hicimos el recorrido de x a y en un paso solo, armamos un único modelo que nos lleva directamente.

Una red neuronal procede “por capas”. O sea, primero construye varios modelos simples, que luego se usan para construir un último modelo que lleva al output final. La figura que acaban de mirar representa una red neuronal de una sola capa. Partiendo de las características x, la primera “capa” construye en este caso tres modelos que, en una segunda etapa, son combinados para producir una predicción del precio.

El siguiente ejemplo puede ayudarlos a comprender mejor esta idea. Es como si, con la finalidad de predecir el precio de un departamento, les diésemos todas las variables a tres asistentes y le pidiésemos a cada uno de ellos que arme un modelo para medir la calidad, el tamaño y la seguridad del departamento. Todos reciben la misma información, pero el primero se focalizará en los detalles de terminación, si tiene piscina, etc. El segundo, en los metros cuadrados, la cantidad de baños, etc., y el tercero, en si hay personal de seguridad en el edificio, si la cuadra está bien iluminada, etc. Es importante ver que los tres asistentes usan la misma información, tal vez ponderando de manera diferente distintas características del departamento. Ya terminada su tarea, los tres empleados vuelven con una suerte de índice de calidad, tamaño y seguridad, y entre los tres arman un predictor para el precio.

Una red neuronal funciona de una manera similar, pero de una forma muchísimo más flexible e interactiva. E incluso la instrucción de base ya no sería “armen índices de calidad, tamaño y seguridad”, sino “armen tres índices, como les parezca, que mejoren la capacidad predictiva del modelo”.

Fíjense que una misma variable puede cumplir un rol en uno de los “submodelos” de la primera capa y otro en el segundo. Por ejemplo, que haya piscina se relaciona tanto con la calidad del departamento como con el tamaño (es raro que un departamento de un ambiente esté en un edificio con piscina).

El gráfico de la figura 7.3 (y la historia que les conté) ilustra una red con “una sola capa escondida”, pero podría tener muchas capas intermedias. La palabra “profundidad” en deep learning tiene que ver, justamente, con la cantidad de capas (más profunda resultará cuantas más capas existan).

La revolución de deep learning es de naturaleza estadística, matemática y computacional. Desde un punto de vista estadístico, se basa en ideas que terminaron de fraguar hacia fines de la década del ochenta. Matemáticamente, el problema es bravo, y se relaciona con garantizar que este ejército de modelos simples, puesto a interactuar, pueda representar relaciones tan complejas como uno quiera.

Pero computacionalmente el problema es dantesco, de tan endemoniado, y ahí es donde hubo notables avances en los últimos veinte años. Un modelo simple está determinado por sus parámetros. Modelos más complejos requieren más parámetros. La naturaleza “neuronal” de deep learning hace que, aun con cantidades no especialmente colosales de datos, la cantidad de parámetros sea exorbitante. Por ejemplo, los modernos modelos de MLWP para predecir meteorología (que mencionamos el capítulo 6 y que, en términos relativos, son chicos) usan unos 37 millones de parámetros. La arquitectura GPT-3.5, que impulsa a ChatGPT, tiene 175.000 millones de parámetros.

Si el modelito de regresión que vimos antes tiene que decidir sobre dos parámetros (los que determinan dónde poner la recta en relación con los datos), el problema de elegir los valores de parámetros que se cuentan de a millones es pasmoso, matemática y computacionalmente. Una inmersión en las ideas de deep learning requiere un enorme esfuerzo intelectual (en el apéndice les propongo algunas lecturas). Temo que llega un nuevo spoiler: hay varias capas de lectura sobre deep learning, y ninguna forma “suave” de pasar de una a otra. Una es accesible y se parece a lo poquito que hemos dicho en estas páginas. La segunda capa requiere una formación sólida en álgebra, probabilidad y algo de computación. El acceso a otras capas es mucho más demandante y tentativo, ya que es un campo de análisis que se mueve a gran velocidad.

Y acá es donde los dejo preguntar: “¿Y de dónde viene lo de las neuronas?”. Tiene que ver con que de alguna manera estos modelos replican el modo de funcionamiento del cerebro. Cada unidad de la red (los círculos, en nuestra figura 7.3) es una suerte de “neurona” que recibe información de muchas fuentes y la retiene o la pasa procesada a otro nodo hasta llegar al output.

Deep learning es el motor detrás de la revolución del ChatGPT y la IA, un resultado conjunto de la estadística, la revolución de datos masivos, la matemática y la computación. Obvio que no exento de críticas. Una es su naturaleza de “caja negra”, es decir, el método aprovecha todo tipo de libertad para predecir mejor, y en pos de ese objetivo, recurre a diversas artimañas matemático-computacionales, sin que el modelo resultante tenga una forma simple y elegante como la de una regresión lineal.

Como adelantamos, no se puede juzgar una herramienta estadística sin hacer explícito su propósito. En defensa de deep learning, señalo que es una herramienta fundamentalmente predictiva, no intenta explicar sino predecir. Criticarla por su “oscuridad” es como despreciar un Lamborghini porque no sirve para que una familia con cuatro hijos y dos perros se vaya de vacaciones a Santa Teresita.

Yo hago ravioles, ella hace ravioles: vecinos cercanos

Y para terminar nuestro recorrido por estas "máquinas de sacar promedios", visitemos una que permite lidiar con una gran variedad de circunstancias y que procede de una forma muy intuitiva, casi de fuerza bruta.

A juzgar por el grupo de WhatsApp que comparto con mis vecinos, somos una auténtica bolsa de gatos, máxime en un país donde, desde la perspectiva de las preferencias políticas y sociales, una pregunta como "¿cuánto da 2 + 2?" tiene veinte respuestas distintas en un grupo de quince miembros. Pero desde otros puntos de vista, la cantidad de afinidades con mis vecinos es notable. Muchos nos mudamos al barrio en una época de recambio generacional: éramos parejas jóvenes huyendo del centro, que aprovechamos que los mayores (que también se habían mudado sincrónicamente en alguna generación pasada) se "achicaban" cuando sus hijos se independizaron. Así crecimos, y ahora somos los "viejos", con nuestros hijos postadolescentes en las gateras de la independencia.

De hecho, si alguien quisiera adivinar la edad de cualquier vecino adulto de la cuadra, el promedio de los restantes no es un mal predictor, por las razones ya mencionadas. Y las similitudes apenas comienzan con la edad. En términos relativos, tenemos autos parecidos, nos vestimos de forma similar, vacacionamos en más o menos los mismos lugares, entre muchos etcéteras. Vuelvo sobre lo de "en términos relativos": tengo un auto de distinto color, marca y modelo que el de mi vecino de enfrente, pero muy diferente de los que tienen los vecinos de un lujoso country y de los de un barrio marginal.

Esta idea de "preguntarles a los vecinos" es un motor fundamental de los enormes logros predictivos del aprendizaje automático y la IA en épocas de big data: el método de KNN ("k vecinos más cercanos"), una idea que explica éxitos predictivos que van desde

cómo detectar un tumor en una imagen hasta por qué la fotografía social (la de eventos cotidianos, que ahora inundan las redes sociales) mejoró abismalmente en los últimos veinte años.

Supongamos que un vecino nuevo se muda a la cuadra. De él sabemos varias cosas, pero no su edad, y vamos a predecirla tomando la edad promedio de los otros vecinos. La calidad de la predicción dependerá de varios factores, a saber:

de cuánto se parezcan entre sí los vecinos,

de a cuántos vecinos les preguntemos,

de qué entendamos por “vecino”.

Como sospecharán, estas cuestiones no son independientes.

Si por razones de confiabilidad quisiésemos resolver el problema predictivo consultando a, digamos, 100 vecinos, muy posiblemente nos enfrentaríamos a que varios de los vecinos de la cuadra estén trabajando cuando intentamos hacer la pregunta y, entonces, para satisfacer este requisito, tengamos que recurrir a vecinos de otras cuadras. Así, corremos el riesgo de empezar a mezclar peras con manzanas si, a fin de mantener constante la cantidad de encuestados, tenemos que alejarnos del barrio. Por otra parte, si la condición fuese “preguntales a los vecinos, pero no te excedas de los que viven en la misma cuadra”, seguramente tendríamos que restringir la encuesta a unos pocos; parecidos pero pocos. En fin, nos hemos topado, una vez más, con un problema “de manta corta”: si queremos mantener fijo el número de observaciones, tenemos que sacrificar “vecindad”, y si queremos dejar fija la idea de vecindad, tendremos que sacrificar el número de encuestados.

El método KNN resuelve este dilema de una manera muy directa. Por ejemplo, si se trata de pronosticar la edad de un nuevo vecino, si $k = 10$, el problema se resuelve tomando la edad promedio de los 10 vecinos que más cerca vivan de la casa del flamante habitante del barrio.

Entonces, el método requiere ponerse de acuerdo respecto de dos cosas. Una es a cuántos vecinos preguntar, es decir, elegir "k", y la otra, qué quiere decir lejos o cerca. Con relación a la primera, lo obvio parece ser "cuanto más, mejor"; pero no, porque, como dijimos, agrandar k implica preguntarle a gente que está más lejos. Pateemos este problema para más adelante, ténganme paciencia.

En cuanto a la distancia, es matemáticamente complejo, pero intuitivamente fácil de entender cómo lidiar con ella. Si se tratase de la distancia "clásica", es cuestión de tomar una regla (o una cinta métrica) y medir a cuántos metros está cada vecino con respecto a la casa del vecino cuya edad se quiere pronosticar, así de fácil e indiscutible. Pero convengamos en que hay otras formas de pensar que alguien está "cerca", independientemente (o no) de lo que diga una cinta métrica. De hecho, la lógica de Tinder (la popular app de citas) es una celebración del hecho de pensar la "distancia" entre dos personas de una manera más creativa.

Por ejemplo, en el caso de Tinder, podríamos pensar en un largo cuestionario de características: equipo de fútbol, película favorita, dulce o salado, de grasa o de manteca, Luis Miguel sí o no, batata o membrillo, playa o montaña, con cientos (¡¿por qué no miles?!) de etcéteras y tomar como distancia algo por el estilo de la cantidad de coincidencias: a más coincidencias, más cerca. Palabras más, palabras menos, el algoritmo de Tinder explota esta idea para "maridar" a dos personas sobre la base de esta noción de cercanía.

Ahora volvamos al problema de predecir la edad del vecino nuevo. Si no dispusiésemos de esa información sobre el recién llegado, pero sí de otra –por ejemplo, si tiene auto, si tiene estudios universitarios, si es extranjero, si está en pareja, etc.–, más que preguntarles a los "k vecinos más cercanos" de acuerdo con la cinta métrica, podríamos consultar a los más cercanos apelando a esta noción de afinidad. Por ejemplo, si el nuevo vecino tiene un Volkswagen, es abogado, nació en Uruguay y está casado, la edad

de otro vecino mío, Marcelo (está a dos cuadras de casa), que es contador, tiene un Peugeot, se mudó desde Chile y vive con su esposa tal vez sea un mejor predictor de la edad del vecino nuevo que la de quien vive en la casa de al lado de él, tiene un Rambler Ambassador, es porteño “nyc” (nacido y criado) y nunca estuvo en pareja. En este caso, estar cerca en la métrica de las características es más interesante que estar en la métrica de la distancia “física”.

Lo interesante de esta cuestión (y una gran “avivada” de machine learning) es que al resolver un problema resolvemos el otro. Les dije que en esto de preguntar a los vecinos había que decidir a cuántos y cómo computar la distancia. Pero fíjense que, si somos capaces de armar una distancia muy “rica” en características, ya no es tan relevante entrevistar a demasiada gente. Digo: si, puestos a adivinar mi edad, se la preguntasen a alguien que es de Boca, le gusta el café, toca la guitarra, prefiere a Deep Purple por sobre Taylor Swift, el café con leche más que el latte, sabe qué quiere decir “Alcoyana-Alcoyana”, tiene canas, presbicia y toma un antihipertensivo todas las mañanas, es muy posible que la respuesta se parezca a la verdad: no soy yo, sino alguien que se me parece mucho, está cerca de mí, no físicamente sino en relación con varias características observables.

Big data aporta un importante ingrediente a esta “salsa” de cercanía: la proliferación de datos trae a la vez más personas y también más características. Con suficientes datos y eligiendo inteligentemente la idea de distancia, no hace falta hablar con muchos vecinos porque, en el mar de big data, hay más chances de encontrar a alguien con características parecidas a las de nuestro vecino, cuya edad puede servir para predecir la edad que no sabemos.

En tratamiento de imágenes, la idea de vecinos cercanos es decisiva. Por ejemplo, si se trata de “pintar” un píxel de un color que no sabemos, la lógica de vecinos cercanos permite decidir qué color usar sobre la base del color observado de los píxeles vecinos. Misma cosa en reconocimiento de tumores en tomografías y muchísimas aplicaciones similares en medicina.

“Vecinos cercanos” es una idea crucial en pronósticos, porque, más allá de las complejidades matemáticas que sustentan sus

propiedades, es fácil de entender, ya que a la larga se basa en predecir usando el valor más cercano disponible en una “métrica” (distancia) apropiada. En el mismo sentido del mítico “Yo hago puchero, ella hace puchero; yo hago ravioles, ella hace ravioles” de la entrañable China Zorrilla en una recordada escena de Esperando la carroza, refiriéndose a su vecina que la copiaba en todo.

Volver del futuro

En tiempos de IA, big data y machine learning, aparecen como hongos los cursos que prometen enseñar herramientas, algunos honestos y otros sospechosos. Este capítulo revisa algunas herramientas predictivas (promedios, regresión, deep learning, vecinos cercanos) no para aprender a usarlas (para eso están los cursos, o Roccatagliata, que en esto se mueve como pez en el agua), sino para invitarlos a reflexionar acerca de qué se puede esperar (y qué no) de la tecnología en la cuestión de los pronósticos.

El promedio es el predictor más simple y, en un contexto general, el mejor.

Una idea poderosa es que, como herramienta predictiva, uno siempre puede ganar si logra “condicionar el promedio”, o sea, afinar el promedio para referir a una circunstancia puntual que nos interesa.

El análisis de regresión es una herramienta fundamental para “retunear” promedios y condicionarlos a eventos más particulares. Así, pasamos de “el precio promedio de un departamento” (un promedio simple) a “el precio promedio de un departamento de 53 m^2, de dos ambientes, en el barrio de Núñez, con seguridad privada” (un promedio condicional).

Deep learning, el verdadero motor detrás de ChatGPT y de mucho

de la IA, es una versión biónica del análisis de regresión: un ejército de modelos relativamente simples que, puestos a interactuar, permiten aproximar realidades hipercomplejas, salteando el monstruo de la “maldición de la dimensionalidad”, que afecta al análisis de regresión cuando aumenta la complejidad.

“Vecinos cercanos” es un mecanismo computacional de “fuerza bruta” que predice por “creación de pares”, fundamental en reconocimiento de imágenes, apps de citas y redes sociales.

8. Predecir es muy difícil, en particular, el futuro: finanzas, economía y series temporales

“Predecir es muy difícil, especialmente el futuro”, dicen que dijo Niels Bohr, que ya ha aportado una frase brillante al capítulo 1 de este libro. En el intento, muchos creen poder explicarlo todo: que salió colorado el 32, que Argentina ganó el Mundial, que vino el covid-19, que se fue, que lo del Brexit, que el papa Francisco. En nuestro bendito país, es casi un deporte olímpico esto de racionalizar con el diario de mañana, todos tenemos un amigo o amiga que lo practica habitualmente.

Cuando mi hijo tenía unos 8 años le regalamos un juego llamado “Dados de Historias”, una cajita hermosa con ocho dados que tenían una imagen distinta en cada una de sus caras. El juego consistía en lanzarlos y armar una historia que conectara todas las imágenes que habían salido. El objetivo era estimular la imaginación y la narrativa. “¡No hay respuestas incorrectas!”, decían las instrucciones.

Desmadejar el pasado es una parte esencial de la actividad predictiva y, en un sentido amplio, de eso se trata la ciencia. Pero en muchas ocasiones esto de “predecir para atrás” es un burdo ejercicio literario, consistente en poder contar una historia coherente que ata eventos observados, como en el jueguito de mi hijo, pero en el que casi todas las respuestas son incorrectas.

El estudio de cosas que ocurren en el tiempo es tal vez la parte más esotérica y crucial de la actividad predictiva. De los electrocardiogramas a la Bolsa, de las curvas de crecimiento de los niños que usan los pediatras a la evolución de la pobreza, mirar este tipo de gráficos tiene tanto de científico como de irresponsable. Porque el que quiere encontrar, a la larga, encuentra, como los

“Dados de Historias”, pero no necesariamente gana.

Este capítulo es una introducción informal a algunas ideas fundamentales del análisis de series temporales: un conjunto de métodos que permiten predecir sobre la base de encontrar patrones en el pasado y que pertenecen al corazón de disciplinas como la meteorología, las finanzas, la economía, la física o la sismología.

El tema del capítulo es la tensión existente entre lo predecible y lo impredecible. Veremos que hay algo paradójico en decir que algo no se puede predecir. O, de manera similar, que aquello que es impredecible tiene la pésima costumbre de disfrazarse de predecible.

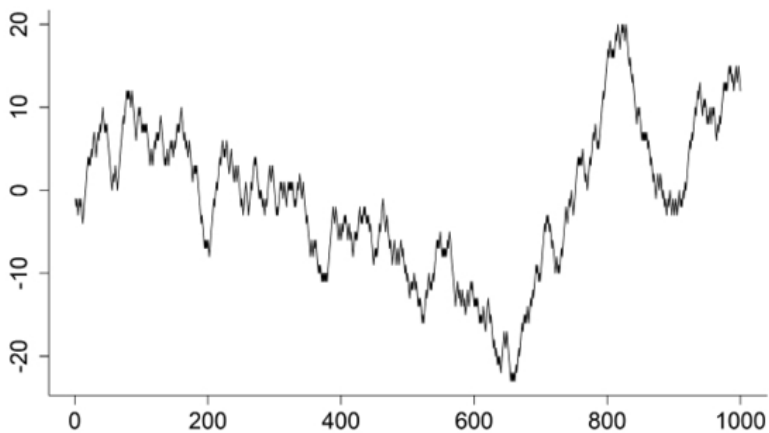
Así que ahí iremos, a recorrer los barrios bajos de las tendencias, los ciclos y los shocks. Disfrácense de detectives. Según prefieran, pueden elegir piloto, sombrero, gafas y lupa, o jeans, zapatillas y remera de los Ramones. Pero les advierto que, al que lo descubren, pierde. Porque los datos son muy buenos embaucando giles.

Pareidolia y soledad: esos esotéricos gráficos financieros

"Pareidolia" es, según Wikipedia, "un fenómeno psicológico en el cual un estímulo vago y aleatorio (habitualmente una imagen) se percibe como una forma reconocible, debido a un sesgo perceptivo". En criollo, el que quiere ver en una nube la cara de Carlos Gardel (o la de nuestra ya asidua Taylor Swift), a la larga, la va a ver. Y en las cuestiones predictivas, esto es un problema a veces severo.

Practiquemos un poquito con la siguiente figura:

Figura 8.1



Se trata del gráfico de una serie de algo (que no voy a revelar todavía) medida durante mil días, empezando desde el 0. Juguemos al analista y describamos qué se ve. Desde el primer día hasta

aproximadamente el 650, se aprecia una clara tendencia a la baja. Errática, con algunas caídas abruptas intermedias y rápida recuperación; pero, sin lugar a duda, descendente en el período mencionado. La situación se revierte alrededor del día 650, cuando comienza una notoria tendencia en sentido contrario, que persiste hasta pasado el día 800, cuando de nuevo comienza otro ciclo a la baja. En los últimos lapsos, se percibe una marcada recuperación.

Ahora les pido que, con una mano en el corazón, piensen lo siguiente:

Concéntrense en los primeros 200 valores, párense en el día 200 y, sin los datos del futuro, sean honestos: ¿realmente habrían estado dispuestos a decir que estábamos en “una clara tendencia a la baja” en ese momento?

Mirando solo los primeros 600 valores, ¿habrían estado dispuestos a decir que a partir del día 650 empezaría una clara tendencia alcista?

Y así, podríamos seguir ad infinitum. Una cosa es describir el pasado y otra, predecir el futuro.

Un condimento muchas veces inevitable de la práctica de la predicción es una familia de gráficos intimidantes como el que les mostré y con el que jugamos un rato (no insistan, todavía no pienso decirles de qué se trata). Y gráficos como este muestran cómo algo se mueve en el tiempo, desde la frecuencia cardíaca hasta los precios de las acciones, pasando por el dólar y la temperatura diaria. Y sí, contribuye mucho a la mística que los demás perciban que uno entiende esos gráficos que, además, dotan a la práctica de pronosticar de una oportuna pátina seudocientífica, más o menos como los verdes enzolves, la cataforesis y, ahora, el ácido hialurónico, palabras frecuentes en publicidades televisivas, que

muy pocos saben qué significan pero suenan a cosa seria.

Agreguemos un poco de terminología predictiva. A una secuencia de datos ordenados en el tiempo se la llama serie temporal, de tiempo o cronológica.

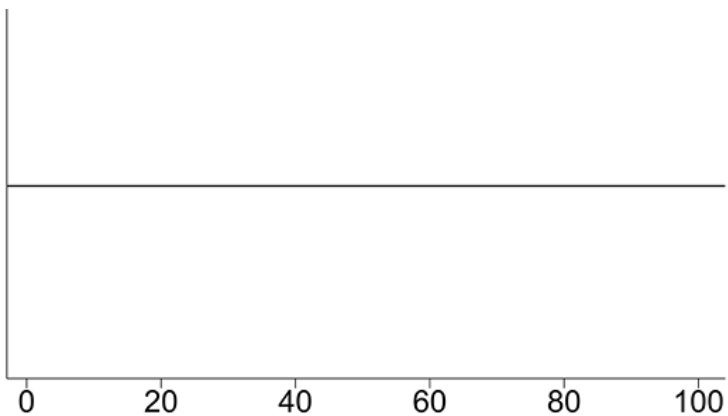
La pregunta clave es si, más allá de la pareidolia, es posible usar la información de estos gráficos para la predicción. La respuesta es esperablemente afirmativa, pero requiere de ciertos cuidados.

Series estacionarias y la predecibilidad de lo impredecible

Las series temporales más fáciles de predecir son aquellas que son estacionarias. Primero, no confundir con estacionales, que sugiere que anda dando vueltas un fenómeno estacional, como que en verano se vende más helado que en invierno. “Estacionaria” quiere decir lo que el nombre sugiere: estable, que se queda quieta en algún sentido, que pasa el tiempo y hay algo que no se mueve.

Obvio que la serie más estacionaria de todas es la de nuestra figura 8.2. Aquí llega:

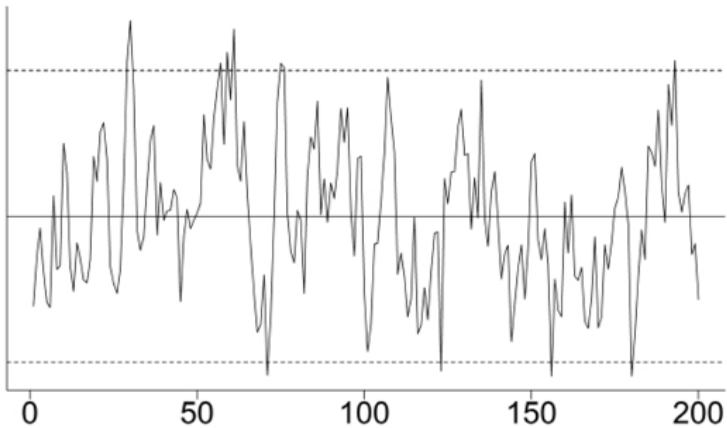
Figura 8.2



Esta serie podría ser la temperatura de una habitación donde no hay ningún cambio en el hábitat, o el precio del dólar cuando en la Argentina teníamos convertibilidad. Esta serie es aburrida, pero convengamos en que es la más fácil de predecir. Acá la predicción es un ejercicio tan trivial como inútil.

Bien. Ahora déjenme mostrarles en la figura 8.3 un caso de serie estacionaria para nada trivial.

Figura 8.3



Tienen derecho a preguntar: "¿Estacionaria? ¿No habíamos quedado en que quería decir 'estable'?". Y he aquí el asunto principal de esta sección: estable no necesariamente equivale a constante. La idea de estacionariedad es que, en cierto sentido, la serie se mueve, pero en otro, no, es estable, como en la figura de recién, la 8.2. Pero no nos vayamos de la figura 8.3; nos toca mirar con detalle (ustedes disculparán, pero todavía me niego a darles un ejemplo de qué fenómeno representa esta serie; ya vendrá, a su tiempo) y pensar qué se mueve y qué no. Y, a larga, qué parte del comportamiento de esta serie es predecible y cuál no.

Primero, si bien erráticamente, la serie parece oscilar en torno a 0. Ese es su promedio, que marqué con una línea horizontal. Fíjense que la serie misma anda por cualquier lado, pero suele pasar por ese promedio, como los hijos ya casados que, aunque no viven con sus padres, los visitan seguido. Es más, la serie no se va sistemáticamente para arriba ni para abajo. Entonces, como les anticipaba, este tipo de serie tiene un promedio constante. Y me

permito un poco de énfasis: ahí tienen la primera cosa que se quedó estable en este tipo de serie, su promedio. Además, aparece el primer sentido en el cual estas series son predecibles: si ustedes apuestan al promedio (0, en este caso), nunca aciertan, pero por lo general no les va tan mal con ese pronóstico.

Pero la cosa no se agota ahí. Hay otro sentido en el cual la serie es estable y, además, predecible. Concentrémonos en los primeros 100 períodos. Ya sabemos que el promedio es estable, pero ahora les pido que hagan foco en el rango de variabilidad de la serie, en el eje vertical. Hay formas rigurosas de hacerlo; en la figura 8.3 les muestro una banda horizontal (entre las líneas punteadas), que incluye a la mayoría de los valores. Y acá viene lo importante: aproximadamente, esa banda permanece inalterada si miramos más datos, si lo hacemos para otros períodos, más cortos o largos. Entonces, lo que aprendimos es que para esta serie no solo el promedio es estable, lo mismo pasa con su varianza.

Por lo tanto, las series estacionarias son predecibles en un segundo sentido, su rango de variabilidad. Si la predicción es ahora un poco más permisiva y si en vez de pronosticar un número (el promedio, por ejemplo), predecimos una banda a su alrededor que contiene a la mayoría de los valores, resulta que la predicción es muy confiable. Y una parte central del análisis de series temporales reside en construir predicciones puntuales y "bandas de confianza" que garanticen cierta cuota de esa misma confiabilidad.

Ya sabemos que una serie estacionaria tiene promedio y también varianza constante. Hemos dejado para el final el elemento más difícil y el más útil para la predicción: que las series estacionarias tienen estructura estable. Fíjense, por ejemplo, una característica más difícil de ver: cuando la serie está en la zona de arriba, tiende a quedarse un ratito por allí, y lo mismo por abajo. Y esa es otra cuestión que hace a la estabilidad. Si recapitulamos, en este caso tenemos:

promedio estable,

varianza estable (la banda que contiene a las observaciones se queda quieta),

estructura estable (la serie tiende a quedarse un ratito donde estaba, ya sea por arriba o por abajo).

Entonces, las series estacionarias tienen muchas características predecibles. Y para fijar estas ideas llega la figura 8.4. Veamos sus cinco series.

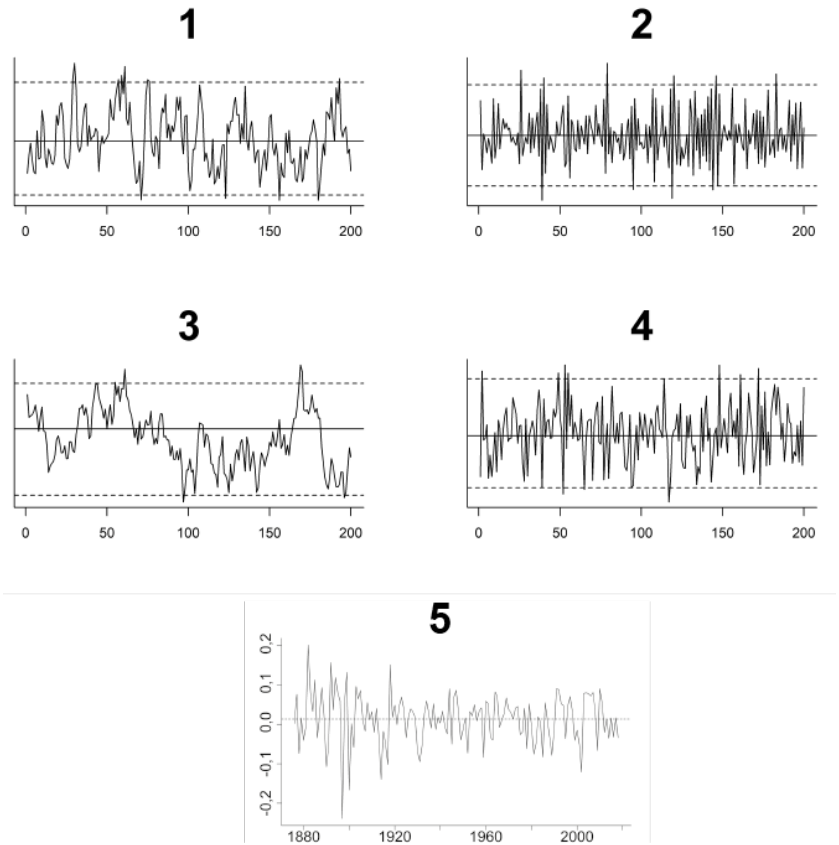
La primera es la que ya discutimos con detalle. Ahora les pido que miren las otras cuatro y piensen en lo siguiente:

¿cuál de las cuatro les parece que es la más predecible?,

¿y la menos predecible?;

para cada una de ellas, ¿cuáles son características predecibles que se pueden detectar a simple vista?

Figura 8.4



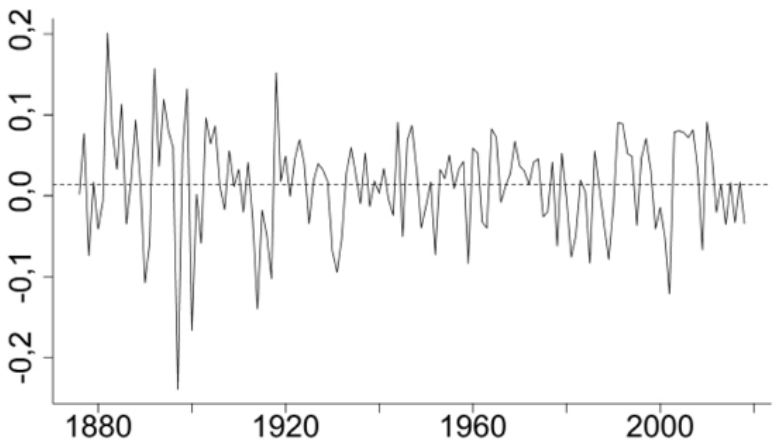
Bueno, ahí vamos de nuevo. Por favor, concéntrense en la segunda serie. Tiene promedio y varianza constantes (la banda se queda quieta). Pero en esta serie notamos una característica muy importante: si en un período está arriba del promedio, inmediatamente se va por debajo. Por ende, esta serie, que parece muy errática, tiene un comportamiento hiperpredecible. Y al mencionarlo se me ocurre que la temperatura dentro de mi auto parece seguir ese patrón. Sí, mi auto tiene un “temporizador automático” que enfría cuando hace calor (y lo contrario si hace frío). No debe de ser muy bueno que digamos, porque cuando siento frío, de repente enciende la calefacción al máximo (y lo contrario cuando empiezo a sentir calor)… calculo que querrá pasteurizarme. Es decir, en mi auto, la temperatura oscila muy rápidamente alrededor de la que yo haya fijado en el termostato. No bien los que

nos dedicamos a esto olfateamos este patrón, respiramos aliviados: es una situación que nos resulta fácil. A no perder la calma, que nadie se da cuenta de esto en un primer intento. Lidiar con series temporales requiere mucha preparación técnica y mucho ojo. Lo mismo pasaría si mirasen una radiografía. Solo el ojo experto y la formación del médico llegan a ver en el caos (y a descifrarlo).

La tercera serie también tiene promedio y varianza constante. Y, si se fijan bien, eso de "permanecer fuera de la media por un tiempo" está más acentuado en esta serie que en la primera. También es fácil hacer predicciones al respecto, porque si está arriba de la media tiende a quedarse por ahí, y lo mismo si está debajo.

Estas estructuras "autorregresivas" (empiezo a llamarlas así) son clave en el diseño de pronósticos económicos. La quinta serie, que volvemos a mirar a continuación, en la figura 8.5, es un gráfico de las tasas anuales de crecimiento del PBI argentino, en términos reales, desde 1875 hasta 2018 (describo en detalle la fuente en el material complementario online; QR disponible en el apéndice).

Figura 8.5

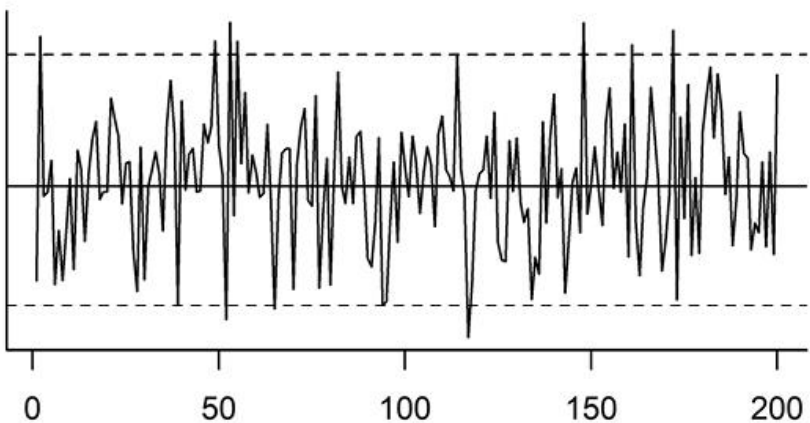


La línea punteada horizontal es el promedio (un 1375%) y, como es de esperar, conviven años de crecimiento espectacular (el 20% en 1882) con crisis catastróficas (un -23% en 1897). Los jóvenes reconocerán la crisis del bienio 2001-2002 o los últimos períodos de crecimiento nulo o negativo. Un análisis muy detallado de esta serie mostraría que es apenas predecible, con similitudes a las características de la tercera serie de la figura 8.4: una leve tendencia a permanecer donde está, pero dominada por mucha erraticidad. Acertarle al tipo modelo que está detrás de una serie estacionaria es una tarea compleja, en la que los avances recientes en machine learning tienen mucho para aportar.

Una alarma en mis oídos: ruido blanco

Y ahora les propongo pasar a nuestra cuarta serie, que vemos ampliada en la figura 8.6, el gráfico más impredecible del bloque que les presentaba antes.

Figura 8.6



Tiene promedio estable, que es "visitado" recurrentemente, un rango de variabilidad también estable y… casi que ya está. A diferencia de las que la acompañaban, no parece haber otro patrón o estructura de correlación que podamos aprovechar para ganarle al promedio como mejor predictor. En este gráfico cada resultado parece completamente independiente del anterior. Si alguien nos regala, digamos, los resultados de los primeros 20 valores, no parece haber patrón alguno discernible; no hay aprendizaje posible. Más allá de servir para calcular el promedio y el rango de variabilidad, los datos no agregan nada.

Este tipo de procesos se conoce en la jerga como ruido blanco, una serie puramente azarosa, en términos de nuestra discusión anterior. Técnicamente, una serie es ruido blanco si surge de una sucesión de "tiros" independientes entre sí, pero con el mismo promedio y rango de variabilidad. El término "independiente" es un poco fuerte –en realidad, lo más exacto es describirlos como "no correlacionados"–, pero, a los efectos de nuestro coloquio escrito, la diferencia no es relevante.

En "Prófugos", Gustavo Cerati dice:

Como un efecto residual

Yo siempre tomaré el desvío

Tus ojos nunca mentirán

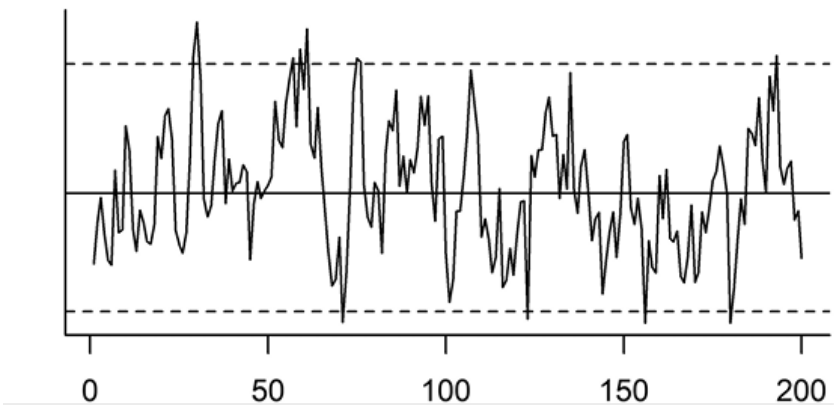
Pero ese ruido blanco

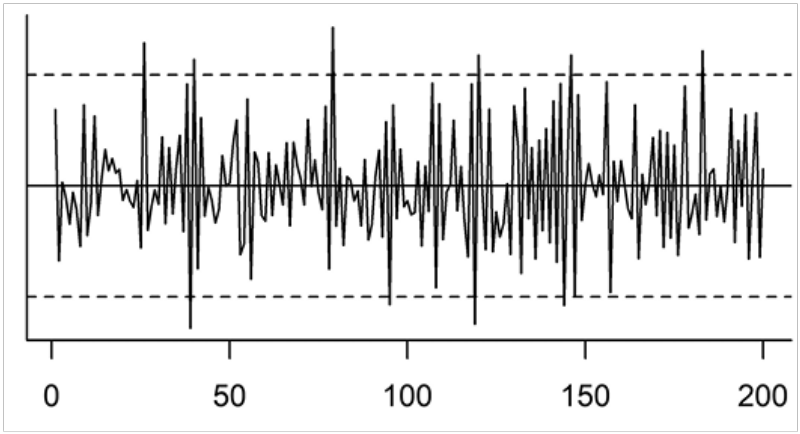
Es una alarma en mis oídos.

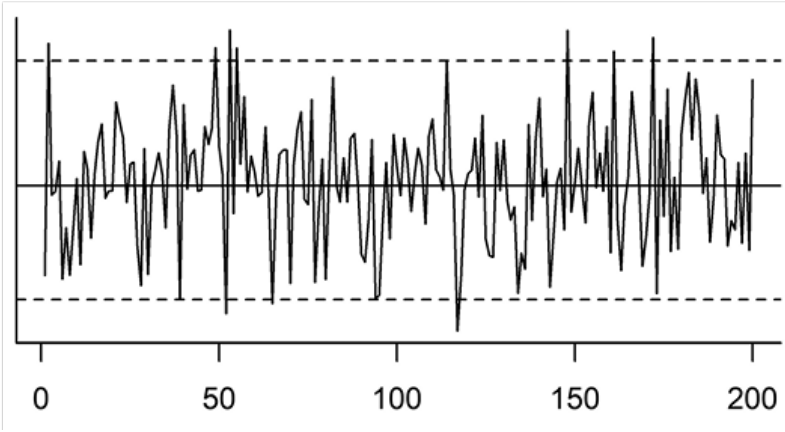
No tengo la menor idea de cuánto sabía Cerati de series temporales, pero la idea de "efecto residual" es central al concepto de ruido blanco. ¿Por qué es tan importante el ruido blanco? Porque verdaderamente es un efecto residual: cuando una serie puede reducirse (o representarse) como ruido blanco, es porque casi no hay nada relevante por predecir, es puro ruido.

¿Es impredecible el ruido blanco? Sí y no. Su promedio es estable y su rango de variabilidad también, y eso lo hace predecible. Más allá de esto, nada para ganar, a diferencia de las demás series, donde el presente dice algo concreto acerca del próximo paso (como en la 1, en que, si en un período está arriba del promedio, es muy probable que en el siguiente esté debajo). Tampoco ocurre como en la serie 2, en que, si en un período está arriba, en el siguiente está por debajo del promedio. Tengamos a la vista de nuevo los tres gráficos, ahora reunidos en la figura 8.7.

Figura 8.7







Fíjense muy bien que ese patrón no ocurre en el ruido blanco,

aunque la vista parezca sugerirnos que sí. Es decir, si saben que en un período está arriba, eso no dice cosa alguna sobre qué pasará con el período siguiente. En un ruido blanco, el tiempo pasa y no hay aprendizaje posible más allá del promedio y la varianza.

¿Y por qué "ruido blanco"? ¿Por qué "ruido" y por qué "blanco"? ¿Alguna vez oyeron hablar de una máquina de ruido blanco o vieron funcionar una? Es un aparato que genera un "ruido neutro" que, a los efectos de dormir, es mejor que el silencio. ¿Cómo? El silencio es, por construcción, la ausencia de ruido, lo que debería ser la panacea para quienes pretenden conciliar el sueño o concentrarse. Entonces, ¿cuál es el problema con el silencio? Bueno, que cualquier mínima disrupción se nota demasiado. Un celular que suena a la distancia, una lapicera que cae, un grillo inoportuno: todo resalta demasiado en un fondo de silencio absoluto, como un puntito negro en una pared blanquísima. Así, muchos, más que en solemnes bibliotecas, preferimos trabajar en algún café (como yo, ahora mismo), porque proporciona un piso de sonido estable y razonable. Enmarcados dentro de un saludable barullo de fondo, los grillos, los celulares y otras distracciones menores no se notan; no es el silencio, pero en un sentido se le parece.

Ahora fíjense cómo, sin darnos cuenta, al describir el sonido de un café (o el de una máquina de ruido blanco) usamos las mismas ideas a las que recurrimos para describir la serie de Gustavo Cerati. El "piso de sonido estable" es el promedio constante: más o menos, el volumen permanece tal cual, pero sin ser 0, como el del silencio. Y que el volumen del ruido de un café oscile, pero se mantenga en un rango razonable es, usando jerga nueva, que la varianza del volumen se quede quieta. Pero la característica esencial del ruido del café es que es ruido puramente azaroso, no hay nada en el barullo de fondo que indique que pasó algo ni que vaya a pasar, y ahí está la gran contribución del ruido blanco: su estable impredecibilidad, la que asiste a insomnes, estudiantes y escritores. Mientras se mantenga en un rango, el ruido del café no me distrae: suena música de fondo (recién ahora me doy cuenta de que no me gusta), mis vecinos de mesa conversan, las tazas se chocan, un bebé llora, se oye el ruido de la máquina de café express, mucho ruido, pero ruido que me concentra, que hace que no me distraiga, que se mueve y se queda quieto a la vez, ruido predecible, que mimetiza

otros que resultarían abruptos y me distraerían. Ruido blanco. Si no tienen a mano una máquina de ruido blanco, métanse en YouTube y escriban esas mismas palabras, “ruido blanco”. O, mejor aún, pongan en YouTube “ruido de cafetería para concentrarse” y van a ver que aparece un sonido de fondo como el que acabo de describir. Es el equivalente sonoro de las plantas artificiales.

¿Y si explicamos por qué “blanco”? Bueno, se relaciona con algunas propiedades de la luz, que exceden lo que tengo autoridad para contar de primera mano. Acá Roccatagliata hace un trabajo razonable, por eso algo les voy a mencionar. Lo que oímos día a día, minuto a minuto, es una mezcla de ondas de distinta frecuencia, de las más agudas a las más graves; dependiendo del tipo de sonido, unas se perciben más que las otras. En el ruido blanco, las ondas están mezcladas en igual proporción, por eso tienen ese “sonido a nada” que, paradójicamente, surge de mezclar todo, como ese líquido indigno que queda en el fondo del tarro de helado después de cuatro horas del almuerzo y tiene un gusto indescifrable. Y acude a la cita el paralelismo que les prometía: lo de “blanco” tiene que ver con que la luz blanca es la mezcla de todos los colores del espectro visible en partes iguales.

Random walks, on the rocks

¿Existen las series temporales realmente impredecibles? El ruido blanco no puede ser una de ellas, porque en dos sentidos (su promedio y su variabilidad) es predecible, no así sus movimientos específicos.

Hay algo tristemente paradójico en que una serie sea impredecible, empezando porque la persistente impredecibilidad es una característica estable y, entonces, predecible. De hecho, las películas de misterio juegan con esa tensión: nunca está claro si el asesino es el tipo encapuchado y lleno de cicatrices o el de atildado aspecto de contador público, con mocasines, chomba Polo y pulóver anudado al cuello. El problema de lo impredecible es que muchas veces se mimetiza muy bien con lo predecible. O no.

Cuando en mis clases de pronósticos les pido a mis alumnos que dibujen una serie impredecible, indefectiblemente la mayoría dibuja algo que se parece a un ruido blanco. No se los pedí, pero es casi seguro que, de haberlo hecho, habrían ido por ese lado. Así que ahora se los pido: intenten dibujar una serie que sea realmente impredecible.

Revisemos un poco más la característica esencial de un ruido blanco: es impredecible en un sentido (la posición de un valor no dice nada acerca de ningún otro, en particular del que le sigue) y predecible en otro dos (el promedio histórico no le pega con exactitud, pero es muy razonable y existe una banda que contiene a la mayoría de los casos).

Entonces, si pescaron la lógica, una serie impredecible debería retener la primera fuente de impredecibilidad de un ruido blanco, pero no los otros dos factores: el promedio en sí mismo debería ser impredecible y no debería haber ninguna banda estable que contenga a sus valores.

Y, como en cualquier capítulo de Criminal Minds, les revelo que el "asesino" ya anduvo dando vueltas por este libro, solo que no se los presenté de esa forma: la serie graficada en la figura 8.1 es realmente impredecible. Peor aún (y aquí cosecho los indicios que fui sembrando acerca de la pareidolia), ya puedo confirmarles que esa furtiva serie se disfraza muy bien de algo predecible. En tren de revelaciones, confieso que la creé con una computadora, siguiendo un procedimiento que vamos a explorar. Empiezo a dar coordenadas.

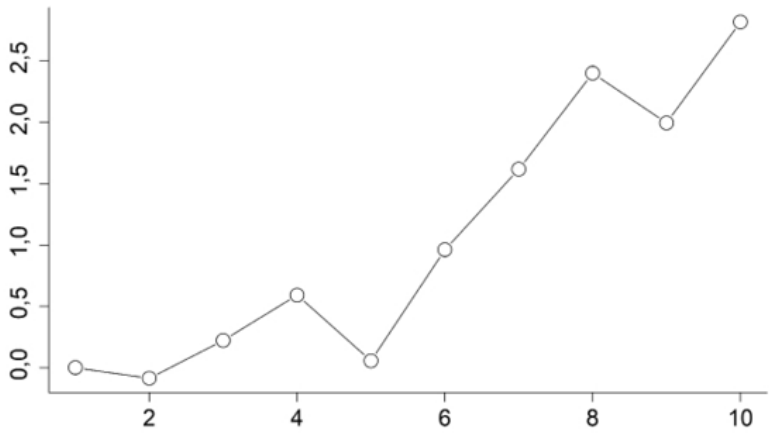
El ejemplo clásico, arquetípico, de procesos impredecibles como este es el paseo al azar o random walk (en inglés). Supongamos que nos paramos a la salida de un bar y que, mirando desde arriba, con ayuda de un dron, perseguimos a uno de los borrachines que de ahí salen. Para simplificar el asunto, supongamos que esta persona se mueve siempre hacia adelante, nunca hacia atrás. Este andar se caracteriza por ser puramente azaroso: el beodo da un paso y, no bien lo dio, olvida lo que hizo y da otro paso, también al azar, y así sucesivamente.

Y acá el profesor en mí tendría que tomarles examen, pero como me sobrepongo y soy bueno, eso no sucederá. Me limito a preguntar: si hiciésemos un dibujo de los pasos (de los pasos, no del recorrido, ojo, ojo) del borracho (paso 1, paso 2, paso 3, etc.), ¿a qué se parecería? Sí, acertaron: a un ruido blanco. Recuerden que un ruido blanco es una serie de valores independientes.

De todos modos, para reconstruir el andar del borracho, tenemos que sumar los pasos. Si empezamos en 0, le sumamos el primer paso, después el segundo, y así sucesivamente, obtenemos lo que muestra la figura 8.8.

Notamos que, empezando en 0, el primer paso fue a la derecha, después hay dos a la izquierda, otro a la derecha, tres a la izquierda…

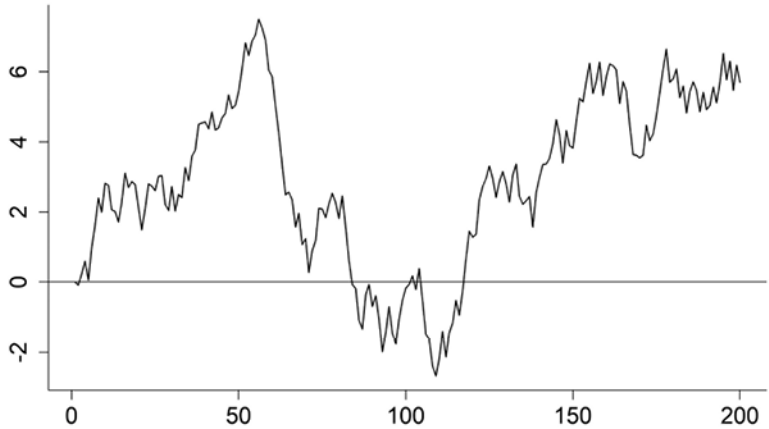
Figura 8.8



¿Y lo impredecible? Nuestro personaje, su andar y el gráfico resultante no tienen pinta de ser demasiado descabellados. Lo suyo, más que de un asesino de serie televisiva, parece cosa de un “extra” cualquiera: una señora que pasa atrás de los protagonistas con el changuito del súper o un vecino que cruza la calle.

Bueno, ahí vamos. Espero haberlos convencido de que la representación de los pasos es como la de un ruido blanco. Y de que el andar del borracho es la suma acumulativa de los pasos al azar. En la figura 8.9 les muestro la suma acumulada de 200 pasos del borracho.

Figura 8.9



Primero, ajústense los lentes (o busquen una lupa) y fíjense que los diez primeros valores de esta serie son los que están en la figura 8.6. En segundo lugar, observen que, en cierto sentido, este gráfico se parece mucho a la figura 8.1. También podrían practicar un poco, como entonces, y hablar de tendencia alcista, bajista, ciclos, etc.

Bienvenidos a la esencia de lo impredecible. El paseo al azar es el asesino que durante un tiempo nos mantuvo engañados y en vilo. Ahora sí tenemos todos los elementos que hacen falta para deschavar su astuta impredecibilidad.

La figura 8.9 no parece sugerir ningún promedio. Peor aún, si construyésemos una banda que contenga a, digamos, la mayoría de los valores de los primeros 20 períodos, veríamos que sería necesario ir agrandando esta banda conforme avance la serie. Entonces, no existe forma de armar una banda de ancho constante que contenga a todos los valores. Si incluso les mostrase 200 valores más, verían que habría que agrandar cualquier banda que contuviese a esos primeros 200. Un espanto. Si no me creen, pueden hacer lo que ya se ha vuelto una sana costumbre en este libro:

creerme porque lo digo yo, que me dedico a esto, soy un vivo bárbaro y por eso escribo libros sobre estadística, ¡habrase visto!,

construir su propio ruido blanco, extender la serie tanto como quieran y ver que efectivamente es así,

demostrarlo matemáticamente.

Para la segunda opción, les dejo online dos líneas de código (en el apéndice, verán QR para llegar). En cuanto a la tercera alternativa, también encontrarán allí alguna recomendación.

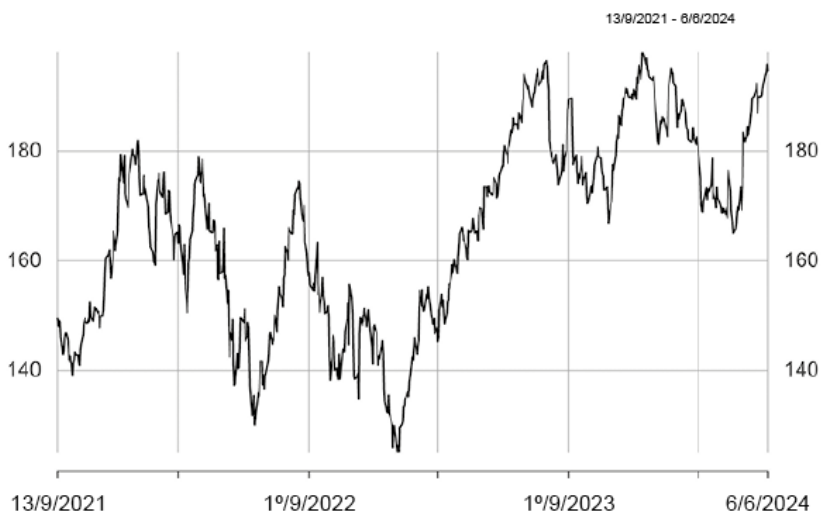
En definitiva, el paseo al azar es virtualmente impredecible, porque, dado que está en un lugar, en el próximo paso puede ir a parar a cualquier otro. Es la auténtica representación de lo impredecible que, muy a nuestro pesar, da letra a cualquier pareidólico, porque, a la larga, le va a mostrar lo que tiene ganas de ver, siempre –indefectiblemente– con el diario de mañana.

Tal vez el sentido más interesante en que el random walk es impredecible es que, si se sabe que quien camina se va a mover, la mejor predicción es que se va a quedar quieto. Para culminar la analogía del borracho, si lo encontramos en alguna parte, su próximo paso será a la derecha o a la izquierda, con idéntica chance y sin relación alguna con los pasos anteriores. Y si decimos que la mejor predicción es que se va a quedar quieto, se debe a que el promedio de la derecha y la izquierda da 0.

El chofer designado de la Bolsa

Todo muy simpático en esto de los paseos al azar y los borrachos, pero… ¿a quién cuernos le importan semejantes minucias en la real realidad? Bien, acá nomás, en la figura 8.10, les muestro una serie de valores del precio de las acciones de Apple.

Figura 8.10

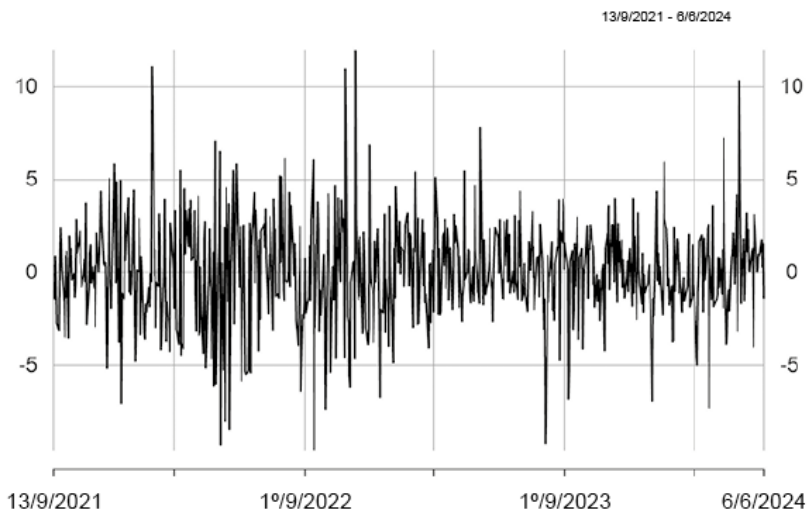


Primero, les pido que se pongan a jugar al analista financiero. Que encuentren tendencias, ciclos, quiebres, cortes y quebradas (como decía Silvio Soldán en Grandes Valores del Tango). Segundo, repitan el ejercicio de ver si esta serie respeta el promedio y si hay una banda estable que contenga a todos sus valores. Y si es un mamífero, de cuello largo, pelaje moteado, con cuernos y que vive

en África… casi seguro es una jirafa. O sea, por lo menos a primera vista, la serie de marras parece seguir un paseo al azar.

De acuerdo con lo que vimos, de ser cierto esto del paseo al azar y la impredecibilidad, deberíamos chequear que los pasos sean efectivamente ruido blanco. Pero ¿qué son los "pasos" en esta historia? Bueno, si los niveles son los precios de las acciones, los pasos son los cambios en los precios. Entonces, aquí la figura 8.11, el gráfico de los cambios en los precios.

Figura 8.11



A ojo de buen cubero, los cambios en los precios tienen aspecto de ruido blanco. Claramente, el ojo es una herramienta bastante imprecisa para evaluar estas hipótesis. En la práctica, se hace alguna prueba estadística. Un ejemplo es el test de Ljung-Box, que, implementado en este caso, sugiere que la hipótesis de ruido blanco es compatible con los datos de los cambios en los precios de las acciones de Apple. Evaluar la hipótesis de ruido blanco requiere un trabajo muy fino (y técnico), que excede por lejos los propósitos de

esta sección.

Si toda esta historia se demuestra cierta, los precios de las acciones son esencialmente impredecibles porque sus cambios son ruido blanco. En finanzas, a la hipótesis de que el mercado impide ganar plata mirando los precios pasados de las acciones se la llama hipótesis de eficiencia de los mercados.

La idea es la siguiente. Si al mirar el pasado pudiésemos predecir que el precio sube, entonces saldríamos a comprar antes de que eso ocurra, para así obtener ganancia. Pero el mercado hace su sucio trabajo, porque gente tan perspicaz como nosotros también se daría cuenta, de modo que ir a comprar acciones baratas tiraría el precio hacia arriba. El mercado hace que la información "la acción está barata" fluya rápidamente, lo que sube el precio y elimina la chance de ventajas. En retrospectiva, la mejor predicción es que el precio de la acción va a mantenerse quieto.

Ahora, ¿vale realmente la hipótesis de eficiencia? Yo creo haber perdido la cuenta de la cantidad de estudiantes de economía y finanzas que usaron este punto como tema de tesis para sus licenciaturas; en algún momento fue una plaga, como ahora los cafés de especialidad y antes las cervecerías artesanales. Como toda ley social o humanística (en oposición a las leyes de las ciencias duras, como la física), es difícil que la hipótesis de eficiencia, tal como la hemos presentado, se cumpla en un sentido estricto; en cambio, caracterizaría una parte importante del comportamiento del precio de las acciones si relajásemos un poco el criterio. La hipótesis de eficiencia es el principal argumento detrás de que los precios de las acciones se muevan como un paseo al azar. Hay parvas de estudios (quizá demasiados) sobre esa hipótesis, que la validan y la refutan con idéntico énfasis. Burton Malkiel escribió un bonito libro a favor de la hipótesis de eficiencia (Un paseo aleatorio por Wall Street). Su némesis es el texto de Andrew Lo y Craig MacKinlay Un paseo no aleatorio por Wall Street (A Non-Random Walk Down Wall Street), todavía no traducido.

Cuidado. La hipótesis de eficiencia afirma no que el precio de las acciones es impredecible, sino que lo es solo cuando se mira el pasado de los precios de las acciones. No se trata de ver un vendedor de humo en cada analista financiero, de ninguna manera.

De la hipótesis de eficiencia (si fuese cierta) no se deduce que al observar otras cosas sea imposible predecir comportamientos financieros; de hecho, la principal actividad de los analistas de esta disciplina consiste en mirar tanto los precios como esas "otras cosas", que incluyen la propia dinámica de las empresas que emiten acciones (el mercado, los balances, entre miles de otras), y sobre la base de estas posiblemente puedan realizar predicciones relevantes (más que "el precio se quedará quieto") y ganar y hacer ganar plata a alguien.

La cuestión central de esta sección es que predecir el comportamiento de los mercados financieros no es tan trivial como sugieren muchos vendedores de humo que andan sueltos por ahí, a quienes les cabe la pregunta "¿y si sabés cómo ganar plata fácil, por qué no la ganás vos directamente?". Para colmo, la pareidolia les da mucha letra a los analistas, siempre con el diario de mañana.

Una nota final sobre cómo se hace esto en la práctica. El análisis de series temporales es el pan y la manteca de los analistas financieros, económicos y de muchísimas otras disciplinas en las que se trata de predecir el futuro mirando correr el pasado, incluidos la meteorología, el marketing, las telecomunicaciones, la epidemiología y la criminología. Por supuesto, el ojo ayuda, pero, a la luz de lo analizado en este apartado, es importante complementar (cuando no, sustituir) la vista con procedimientos matemáticos y estadísticos insensibles a interpretaciones absurdas. El test de Ljung-Box, que usamos más arriba, es un ejemplo.

En esta sección miramos de a una serie por vez, pero es natural conjeturar que podemos aprender mucho y mejorar los pronósticos si contemplamos los comovimientos de varias series. Finalmente, el análisis de series temporales requiere una formación matemática contundente. Cosas "raras" para el lego, como las series de Fourier, los espacios de Hilbert y las ecuaciones en diferencias, sumadas a los complejos gráficos de la práctica, cubren esta parte del análisis de datos de una pátina de magia que espanta tanto como atrae a los ajenos a esta disciplina.

Como cabe esperar, el análisis de series temporales ocupa un lugar prominente en todas las disciplinas en que el paso del tiempo es relevante. Cualquier libro del tema incluye ejemplos que van desde

la evolución de las manchas solares en astrofísica hasta la de la población de conejos en una isla, pasando por la de las temperaturas máximas (elemento clave en el estudio del cambio climático), la actividad sísmica, la cantidad de autos que transitan por un puente o, en lingüística, el estudio de patrones de lenguaje en las redes sociales; por si fuese poco, también es una pieza importante en los sistemas que reconocen canciones o lenguaje oral.

Tuve la fortuna de tener de profe a Paul Newbold, uno de los padres del análisis de series temporales y autor de dos libros fundamentales sobre pronósticos. Con su finísimo humor inglés, un día dijo en clase: “La única forma de hacer plata con el análisis de series temporales es escribir un libro de análisis de series temporales”. La fruta no cae lejos del árbol.

ChatGPT para pronosticar la inflación

Y si de encontrar patrones temporales se trata, ¿cómo le irá a Roccatagliata, todo un experto en esto de ver en los datos? Razonablemente bien, como veremos.

El mítico Frank Zappa ponía a prueba a sus músicos haciéndoles tocar "La Página Negra", tal vez la composición más endiablada de la música moderna, como una visita a YouTube permite verificar. Y ante la actitud pendenciera de ChatGPT y los large language models (LLM) de la IA –que parece que lo pueden todo–, el equivalente económico de lo de Zappa sería ponerlo a predecir la inflación, a ver si tiene uñas de guitarrero. Exactamente eso hicieron Miguel Faria e Castro y Fernando Leibovici, investigadores de la Reserva Federal de Saint Louis, en un muy reciente paper.

La inflación es una bestia compleja y, en la Argentina, por las peores razones, no hace falta aclararlo. Las dificultades tienen que ver con que la inflación refleja las conductas de todos los actores de una economía (consumidores, productores, gobierno, etc.). Peor aún, la predicción de la inflación depende también de las expectativas que todos esos actores tengan de la mismísima inflación.

A la fecha, ChatGPT no es una herramienta adecuada para esta tarea por varias razones; la más importante es que su stock de información disponible se detiene en septiembre de 2021. Por esto, Faria e Castro y Leibovici apelaron a PALM, la versión de Google de los LLM.

Tal vez la principal capacidad (y desafío) de los LLM sea su aptitud para contextualizar: el tipo de respuesta que pueden dar depende de la forma en la que se les formulan las preguntas o prompts. La tan en boga idea de "ingeniería de prompt" refiere a la habilidad de acertarle al contexto adecuado para formular preguntas a alguno de los LLM.

Entonces, la forma como se le indica a PALM que haga pronósticos es crucial. En concreto, luego de experimentar con varias alternativas, Faria e Castro y Leibovici usaron el siguiente prompt para pedirle a PALM que realizara un pronóstico de la inflación en los Estados Unidos:

Supón que estás en el período x. Por favor, dame tu mejor pronóstico de la inflación del IPC (índice de precios al consumidor) ajustada estacionalmente en los Estados Unidos para t, t + 1, t + 2, t + 3, t + 4. Proporciona valores numéricos para estos pronósticos. No utilices ninguna información que no estuviera disponible para ti en x para formular estos pronósticos.

El ejercicio consiste en repetir esta consigna para varios años (x, en el párrafo anterior) y luego contrastarlos con la inflación observada y con la pronosticada por otros medios. El estudio en cuestión obtiene predicciones de PALM y las compara con las de la Encuesta de Pronosticadores Profesionales (SPF, por sus iniciales en inglés), para el período 2019-2023. El SPF es una encuesta que se realiza de forma periódica desde 1968, en la que intervienen 36 expertos que realizan pronósticos de variables económicas como el PBI, el desempleo y la inflación.

Hay malas y buenas noticias. Empecemos con las malas: tanto a los expertos como a PALM les va más o menos mal en predecir la inflación. Lo cual no es ninguna sorpresa, tanto por las dificultades inherentes en predecir comportamientos sistémicos –como mencionamos antes– como porque, a la larga, predecir la inflación en ese período implica haber lidiado con el episodio del covid-19, su duración, su extensión y alcance.

La “buena” (con tantas comillas como sea necesario) es que, en muchos sentidos, a PALM le va mejor que a los expertos. Faria e Castro y Leibovici comparan los guarismos de PALM y SPF y encuentran dos resultados interesantes. Primero que, comparado con la inflación observada, PALM comete menos errores que el SPF. Contextualicemos esta aseveración: a ambos les va relativamente

mal, pero a PALM le va un poquito mejor. Este resultado es sorprendente. La implementación del SPF es costosa y lenta; PALM, por el contrario, tiene costo virtualmente nulo y está disponible casi en tiempo real. Esto habla del potencial de los LLM para estas tareas. Si a los expertos del SPF les dijeran que su trabajo será reemplazado por robots, les picará la oreja mientras el paper de estos autores se disemina y actualiza.

En segundo lugar, la tasa de inflación previa al covid-19 para los Estados Unidos rondaba en un 2% anual. Cuando esta tasa crece luego del episodio de la pandemia, tanto PALM como SPF pronostican un retorno al 2%. Y aquí hay una diferencia importante. PALM predice un retorno mucho más lento que SPF, como ocurrió en la realidad. O sea, SPF se equivoca más que PALM, no solo en los niveles sino también en la dinámica de la inflación de mediano plazo. Este no es un punto menor.

Otra dimensión relevante es verificar si PALM respeta la restricción de usar datos hasta un momento deseado y no se tienta con información posterior para mejorar el pronóstico. A tal efecto, los autores juegan con preguntas artificiales para ver si el LLM "pisa el palito". Por ejemplo, le preguntan cosas como "respondé la siguiente pregunta como si fuese el 1º de enero de 2020: ¿está viva la reina Isabel II?". PALM responde que sí, con prescindencia de que en ese momento (2023) la cuasieterna Lilibeth lleva más de un año muerta. Bien por PALM.

Y por casa… ¿cómo andamos? Antes de que el elefante nos pise, refirámonos a él. A la fecha, ChatGPT es el LLM del pueblo, y es natural ver qué tiene para decir en esta cuestión. También es relevante aclarar que los movimientos bruscos de la inflación en los Estados Unidos estudiados por Faria e Castro y Leibovici son una leve brisa en comparación con el tsunami de los guarismos vernáculos. Entonces, procederemos con cautela.

Con exactamente el mismo prompt al que apelaron Faria e Castro y Leibovici para que PALM hiciera pronósticos, le hemos pedido al ChatGPT que prediga la inflación anual para la Argentina de 2018, con información hasta 2017. ChatGPT es mucho más terco (¿cauto?) que su par de Google. En un principio, se niega rotundamente a producir un pronóstico y, además, se disculpa. Algo

de reingeniería de prompt hace que, luego de una ardua batalla dialéctica, ChatGPT se digne a decir que

basado en la información disponible hasta el 31 de diciembre de 2017, una estimación razonable para la tasa de inflación anual en la Argentina en 2018 podría haber estado en el rango de aproximadamente 20 a 30%. Sin embargo, esta cifra es puramente una estimación basada en la tendencia de la inflación en años anteriores y no debe considerarse como un número concreto o una predicción precisa.

¿Son ChatGPT y sus variantes una competencia seria para los pronosticadores profesionales de la inflación y otros guarismos importantes? En meteorología, sí (lo vimos en el capítulo 6). En ciencias sociales, “depende”, según diríamos los economistas. La respuesta, como cantaba Bob Dylan, está flotando en el viento, que en estos días sopla fuerte.

Volver del futuro

En lo medular, predecir el futuro mirando el pasado depende de que haya algo que permanezca quieto. No recuerdo quién dijo que no hay herramienta más inútil que la que no se usa. Que el análisis de series temporales siga ocupando un lugar central en el quehacer cotidiano de las finanzas, la economía, la meteorología o la física aplicada sugiere que predecir el futuro, aunque muy difícil, no parece ser imposible.

El primer escollo que hemos enfrentado en este capítulo es evitar confundir lo errático con lo impredecible. Las series estacionarias aparecen como ejemplos de comportamientos que, en su erraticidad y bajo una apariencia engañosa, contienen varios aspectos predecibles. Estas series se mueven a lo largo de un promedio estable y en un rango relativamente acotado. Y algunas hasta sugieren algunos patrones interesantes, como que, si están por arriba del promedio, tienden a quedarse un ratito por allí.

El ruido blanco es un tipo de comportamiento en que lo único predecible es su promedio y su rango de variabilidad.

Muchos fenómenos económicos (como la evolución del precio del dólar) o financieros (los precios de las acciones) tienen enormes componentes no estacionarios. Son series que no reconocen un promedio obvio y resulta imposible acotarlas dentro de alguna banda de variabilidad razonable.

El truco es que lo impredecible surge de acumular pasos al azar, como el andar del borracho, como los precios de las acciones que se obtienen de sumar los cambios.

El gran problema con estas cuestiones no estacionarias es que se disfrazan muy bien de cosas estacionarias, y más de un inocente cree ver lo que tiene ganas de ver en su recorrido puramente azaroso, siempre con el tan socorrido diario de mañana.

La verdadera razón por la que no es trivial cómo llenarse de plata comprando y vendiendo acciones, y por qué nadie tiene la menor idea de cuánto va a valer el dólar de acá a un mes, no es porque faltan datos, ni por la impericia de los analistas, sino porque los objetos de las finanzas y la economía están plagados de los componentes no estacionarios que describimos en este capítulo.

ChatGPT y la IA tendrán mucho para aportar a esta cuestión de lo temporal.

9. La vida es juego: estrategia y predecibilidad

¿Y si realmente hay cosas que no se pueden predecir? ¿Y si es cierto, como dijo Oscar Wilde, que “la única función de la profecía es mantener viva la esperanza”? A lo largo de este libro hemos discutido distintos grados de predecibilidad. De los éxitos de los sistemas de recomendaciones a los paseos al azar, de los precios de los departamentos al ruido blanco, desde el mismísimo ChatGPT (que, a la larga, es un mecanismo predictivo dinámico de lenguaje) hasta el PBI argentino.

Como el libro de la historia todavía no se terminó de escribir, creemos que la impredecibilidad de ciertos fenómenos es esencialmente una cuestión de tiempo, que será resuelta por el avance de la ciencia, de los algoritmos, de los datos, de la inteligencia artificial y de la natural. Es entendible el entusiasmo, máxime a la luz de los apabullantes avances que puede mostrar la ciencia, ahora envalentonada por el fenómeno de big data y de la IA.

Este capítulo explora una ruta alternativa, que tímidamente hemos tocado en capítulos anteriores: quizás algunos sucesos son efectivamente de una naturaleza impredecible. En estas circunstancias, la impredecibilidad no tiene que ver con impericia o falta de datos, sino que es una consecuencia inevitable de que ciertos aspectos de la sociedad parecen haber sido diseñados para no poder ser predichos. Acerca de ellos, la predicción más confiable, por paradójico que resulte, es que no se pueden predecir.

El juego de los penales y las finanzas

Con su efectivo sentido del humor, el reconocido economista Juan Carlos de Pablo les contaba a sus alumnos que había escrito un libro llamado Cómo patear penales, en el cual explicaba que el truco consistía en patear para el lado contrario al que se tira el arquero. Y que el éxito del libro había sido tal que se había visto obligado a continuar la saga con Cómo atajar penales, en el que sugiere que los goleros se tiren para el lado al que patea el jugador. No muy lejos de esta chanza estuvieron los consejos que me dio un amigo de mi padre cuando me casé (leer con voz ronca, como la del Coco Basile): "Pibe, 'cuchame: el secreto de la vida es comprar cuando está barato y vender cuando está caro".

La inoperancia de estas sugerencias tiene que ver con la imposibilidad de saber de antemano qué lado elegirán el arquero o el pateador del penal, o cuándo algo está barato o caro, máxime cuando en esta última circunstancia los consejos aplican simétricamente a compradores y vendedores. También hay una cuestión de velocidades. Un futbolista profesional patea a más de 60 km/h, de modo que es imposible que un arquero pueda primero ver hacia dónde sale el disparo y luego tirarse en consecuencia. Y algo similar sucede con las finanzas. Una conocida chanza cuenta que un economista le dice a otro: "Uy, mirá, ¡un billete de cien dólares tirado en el piso!". "Imposible", responde el otro, y siguen su camino. La cuestión es que es muy poco lo que puede durar un billete de cien dólares en el piso sin que nadie lo levante.

Y si de predecir el futuro se trata, los economistas la tienen complicada. Difícilmente esto se deba a las incapacidades intelectuales de los practicantes de la disciplina de Smith, Ricardo y Marx. Miembros conspicuos de las ciencias duras han enfrentado los mismos escollos en relación con la dificultad de lidiar con los fenómenos sociales. "No existe una teoría económica universal… y probablemente no la haya durante nuestras vidas. La razón es

simplemente que la economía es una ciencia demasiado difícil", escribió el matemático y físico John von Neumann –una de las mentes más brillantes del siglo XX– en la introducción a su libro en colaboración con Oskar Morgenstern, en el cual sientan las bases de la teoría de juegos, que estudia el tipo de cuestiones interactivas que complican a las ciencias sociales y también a las biológicas.

En concreto, un tipo de impredecibilidad habitual en la economía y la política surge de las situaciones estratégicas, "esas en las que lo que sucede depende de lo que uno hace, pero también de lo que hace el otro, y de lo que uno y otro opinen de lo que ambos piensan", como en los penales, como en el espionaje, como en los mercados financieros.

La teoría de juegos es un área de la matemática y la economía que estudia estas situaciones estratégicas. Sus padres fundadores incluyen a los ya mencionados Von Neumann y Morgenstern, y también a John Nash, cuya vida funambulesca y trágica fue plasmada en la premiada película Una mente brillante. Y resulta casi paradójico, pero una predicción sólida de esta disciplina es que varias situaciones que ocurren en el marco de una interacción estratégica son esencialmente… impredecibles.

En relación con los penales, la teoría de juegos dice que ningún arquero ni delantero profesional deberían seguir una estrategia obviamente adivinable. Cada vez que enuncio esta conclusión en las charlas que doy para estudiantes secundarios, se desata un auténtico maremoto de reacciones encontradas, cuando no gritos y otros exabruptos. Jamás falla: no hay mejor forma de captar la atención de un adolescente que enfrentarlo a sus propios instintos.

Cuando pregunto si existen estrategias para atajar penales, siempre, sin lugar a dudas, alguien quiere contar alguna anécdota: la de un amigo que ataja en un equipo y que hace esto o lo otro, que mira a los ojos al pateador, después gira a la izquierda, mueve la cintura a la derecha o luego se tira a la derecha también, cual Elvis moderno. También me rememoran episodios como el del "papelito de Chiquito Romero" (sobre el que tiempo después se dijo que más que la descripción de qué hacían los pateadores era una carta de su esposa) o el caso de Jens Lehmann en 2016, a quien, supuestamente, alguien le pasó un papel con indicaciones de hacia

dónde iban a patear los jugadores argentinos. Anécdotas.

Bien, intentemos entender esta cuestión de la imposibilidad de predecir qué va a hacer un arquero (o un delantero) en un penal. Vayamos primero a la teoría. Como aventuramos, la teoría (de juegos) dice que, contra lo que sugieren las anécdotas, ningún arquero ni pateador profesional deberían seguir una estrategia obvia. Pensemos un poco. Supongamos que Raúl Rey es un arquero profesional, meticuloso e inteligente, y que, como tal, mira obsesivamente videos en YouTube para ver qué hacen los delanteros en el momento del penal. Sobre la base de toda esta información, diseña una estrategia, tal vez "tuneada" para jugadores específicos que enfrentará en distintos partidos. No, no revelaremos su estrategia, porque este libro es científico pero no buchón.

Tito Bonfini es un temible y efectivo delantero, no más ni menos astuto que Rey, que también está obsesionado con esto de los penales y apela a un método idéntico al de su colega golero: recoge datos, información, habla con cientos de delanteros y arqueros, y procesa miles de penales por YouTube. Y también diseña una estrategia que, por las razones ya esgrimidas, tampoco será revelada en este libro.

Y si cierta tarde Tito y Raúl se enfrentasen en épica contienda de penales, ¿pondrían en ejecución sus estrategias? Qué momento. Si Raúl sigue una estrategia, debería darse cuenta de que Tito (o cualquier jugador inteligente) la habrá adivinado. Y como Raúl es también inteligente, se da cuenta de que Tito se dio cuenta de que él se dio cuenta de su estrategia. En mis charlas a jóvenes, acá paro y pregunto: "Entonces, si Raúl se da cuenta de que Tito se dio cuenta de su estrategia, ¿qué es lo que va a hacer en el penal?". Y en ese trance casi todo el mundo pisa el palito. Les traslado la pregunta, y, una vez más, pido por favor que la enfrenten. Los espero.

Ahora que retomaron la lectura, les cuento que a casi todo el mundo le sale del alma decir: "Va a cambiar la estrategia". Pero si Tito es realmente inteligente, debería darse cuenta de que Raúl se dio cuenta de que Tito se dio cuenta de que tenía una estrategia, y también de que, si va a cambiarla, Tito se va a dar cuenta. ¿Entonces, la cambia o no la cambia? Y ahí está el lío.

Este delirante juego de palabras y de espejos (se dio cuenta de que se dio cuenta de que el otro se dio cuenta, y así sucesivamente, ad nauseam) forja una suerte de contradicción: cuando juegan dos personas inteligentes, tener una estrategia obvia implica no tenerla. Y esto es lo que sucede en la práctica de los penales y lo que predice la teoría de juegos: los jugadores profesionales no deberían seguir una estrategia trivial.

El enigma de Alan Turing

La película Enigma trata de un episodio crucial de la Segunda Guerra Mundial y sigue la titánica causa de Alan Turing. Si no conocen a Alan Turing, deberían interrumpir la lectura de este libro y lanzarse de lleno a averiguar quién fue. Los ayudo: Turing fue una de las mentes más brillantes de la historia, uno de los padres de la computación y la informática moderna. A sus éxitos científicos y comunitarios (como veremos) se suma una historia trágica. Su carrera terminó abruptamente en 1952, cuando fue procesado y condenado por su homosexualidad (que hasta 1967 siguió penalizada en el Reino Unido). Abatido, en 1954 murió envenenado tras morder una manzana, episodio que dio lugar a la leyenda urbana que se esconde detrás del célebre logo de las computadoras Apple. Recién en 2009 la Corona británica pidió disculpas. El indulto por su "delito" le llegó póstumamente, en 2013, de manos de la reina Isabel II.

Volvamos a Enigma, que les recomiendo. Sí, hay spoiling en lo que queda de esta sección, pero, créanme, si no vieron la película, en poco les va a afectar lo que les cuente. Vamos, el mundo entero vio Titanic, aunque todos sepamos cómo termina. Todo transcurre en la Segunda Guerra Mundial. Los Aliados saben que los alemanes usan un código secreto para comunicarse, basado en la máquina "Enigma", un sistema mecánico de rotores para cifrar y descifrar mensajes. Una vez más, los científicos parecen "tener problemita" con el marketing: como con el GraphCast de los meteorólogos, ponerle "Enigma" a un código secreto es una estrategia que se para en esa peligrosa línea que divide a la genialidad de la vagancia, como cierto producto para el cuidado del pelo que más bien remite a una cantina vasca.

Bien, ya estamos en pleno Enigma. Los Aliados confían en la astucia de Turing para "quebrar" el código del Reich, y ese parece ser el nudo de la película. La duración total del film es de ciento catorce

minutos. ¿En qué momento de la película les parece que Turing descifra el código? Casi todos mis alumnos dicen “al final”. Pero no, la estremecedora escena en que Turing desmenuza el misterio ocurre a la hora y trece minutos (y quince segundos, para mayor precisión), un poco después de la mitad de la película. Cualquier espectador iluso se habrá sentido estafado al ver que “el asesino” del thriller se revela tan temprano. A la hora y dieciocho minutos la película parece llegar a su fin, el equipo de Turing festeja el logro, de fondo suena música heroica y los títulos finales parecen prontos a caer. Sin embargo, en ese momento exacto es como si la película empezase de vuelta, y calculo que ustedes ya notaron por qué.

Como Tito y Raúl, mientras sus colegas festejan efusivamente, Turing se da cuenta de que, si los Aliados mueven las fichas equivocadas en el tablero, los alemanes enseguida van a notar que los Aliados se dieron cuenta… y volvemos al loop infinito y abismal de los penales. Turing internaliza la esencia estratégica de la comunicación y la guerra (y los penales y las finanzas y la economía y la política), y un nuevo enigma comienza en Enigma, para beneplácito de los espectadores. Este mismo carácter estratégico tiñe de impredecibilidad fenómenos en apariencia predecibles.

Y aquí vuelve al ruedo John Nash, quien, como veíamos, en 1950 dio un paso fundamental –que le valió el Premio Nobel de Economía de 1994– cuando demostró matemáticamente que, dadas algunas condiciones generales, estos juegos estratégicos tienen un equilibrio. Es muy difícil definir qué es un equilibrio en estas circunstancias; nos limitaremos a decir que es una estrategia para cada jugador, de modo que ninguno tiene incentivos para apartarse de ella. En los penales, una estrategia que dice “el arquero se tira siempre a la derecha y el pateador lo hace a la derecha” no es una situación de equilibrio, porque el delantero tiene los incentivos para dejar de lado esta regla y patear a la izquierda.

El de los penales es un juego estrictamente competitivo, jerga que se usa para referirse a juegos en los cuales lo que uno tiene para ganar es exactamente lo que el otro tiene para perder. En esta situación simétrica, cualquier estrategia “pura” (siempre a la derecha, siempre a la izquierda) beneficia a uno y perjudica al otro. Una consecuencia del resultado de Nash es que, si bien no existe un

equilibrio “en estrategias puras”, sí existe “en estrategias mixtas”, o sea, en casos en que arquero y delantero proceden como si lanzasen una moneda antes de decidir qué hacer y actuasen en consecuencia. Esto es llamativo. Es como si arqueros y delanteros se pusieran de acuerdo en respetar esta seudoestrategia, lo cual lleva a que nadie gana (ni pierde) sistemáticamente por hacer otra cosa.

Otra forma de abordar este problema es que el arquero piense qué es lo peor que le puede pasar si se tira a la izquierda, al centro o a la derecha y elija qué hacer dentro de este marco decisional; dicho de otro modo, que elija la mejor de las alternativas factibles. Por ejemplo, si se tira siempre a la izquierda, lo peor que le puede hacer el delantero es tirarle siempre a la derecha. Entonces, una estrategia que implique “siempre a la izquierda” no tiene sentido para el arquero, porque le da una ventaja obvia al delantero. Este razonamiento se conoce como minimax: los jugadores eligen la mejor estrategia suponiendo que el contrincante les causará el mayor daño posible ante cualquier cosa que hagan.

En este contexto, el equilibrio de Nash es “en estrategias mixtas” o aleatorizado, de modo que es como si el arquero y el pateador jugasen tirando una moneda, apropiadamente cargada. La estrategia de aleatorizar entre tirarse-patear a la izquierda o a la derecha les conviene a ambos.

Pero volvamos al cauce, a la temática de este libro, los pronósticos. La existencia de un equilibrio en estrategias mixtas predice que el resultado de los penales es fundamentalmente impredecible: está sujeto a ciertas chances, que arqueros y delanteros usan para no perder de manera sistemática. En otros términos, parece ser confiable la predicción de que el resultado de patear un penal es impredecible.

La naturaleza imita al arte y la ciencia, a Maradona

Hasta acá, la voz de la teoría. Ignacio Palacios Huerta es un reconocido economista español, especialista en teoría de juegos y estrategia, que además ha sido asesor (y actual miembro del board) del Athletic Club de Bilbao y ha dedicado buena parte de su energía profesional a entender la dinámica del fútbol competitivo.

En un celebrado trabajo de 2003, Palacios Huerta analizó 1417 penales de las ligas profesionales de Europa, entre 1995 y 2000. Los resultados son contundentes en respaldar lo que dice la teoría. Leamos qué notó.

Ninguna estrategia (patear o tirarse hacia determinado lado) resulta más efectiva que otra.

Nada se aprende de mirar historias de penales. Patear y atajar penales es un evento esencialmente impredecible, su propia lógica hace que no tenga sentido jugar con una estrategia predecible.

En 2023 Palacios Huerta actualizó su investigación, pero focalizando en la performance de Diego Armando Maradona. “Maradona juega como el minimax” es el apropiado título del paper. Luego de un meticuloso escrutinio estadístico obtuvo dos resultados, que reforzaron sus hallazgos previos:

Maradona metió el 85,1% de los penales que pateó; cuando lo hizo

a la izquierda, la tasa de aciertos fue del 84,6% y cuando pateó a la derecha, del 85,7%,

en cualquier momento dado, es imposible predecir si Maradona pateará el próximo penal a la izquierda o a la derecha analizando los penales que pateó con anterioridad.

Como discutimos antes, el teorema minimax predice que en situaciones estratégicas como las de un penal:

la tasa de éxito debería ser igual para todas las estrategias posibles y

los jugadores no seguirán una estrategia estricta (y adivinable por el adversario), sino que la aleatorizarán.

Los resultados de Palacios Huerta sugieren que los jugadores profesionales respetan a rajatabla lo postulado en los resultados de Nash y Von Neumann: no siguen una estrategia adivinable, y son igual de efectivos pateando hacia la izquierda o a la derecha. Maradona imita a la teoría y esta, a él, como la naturaleza al arte, como dijo Oscar Wilde.

No es exagerado decir que una parte sustancial de las acciones sociales está regida por las reglas del comportamiento interactivo y estratégico, como las que estudia la teoría de los juegos. La economía y la ciencia política han dedicado un espacio considerable a estas ideas para entender procesos electorales, de coaliciones entre empresas, de negociación de deuda, de determinación de precios de activos financieros, entre muchos otros temas que, por su estructura lógica, se asemejan a la de patear y atajar penales.

La consecuencia más dramática de estos resultados es que la misma teoría (aquella que la práctica de los penales y muchos eventos de la economía y la política parecen respetar a pies juntillas) se encarga de señalar que es imposible predecir exactamente el resultado de ciertos juegos estratégicos, porque los involucrados aleatorizan sus estrategias y, peor aún, porque los datos de la historia de estos sucesos (disponibles en videos de penales) resultan inútiles para la predicción, como demostró Palacios Huerta en su reciente estudio y como lo hizo con una gran variedad de jugadores profesionales en investigaciones anteriores.

Esta última apreciación pone un freno a lo que es esperable aprender de los datos y al sobreentusiasmo que muchos tienen en relación con el fenómeno de big data y la IA, cuando creen que una mezcla de datos y algoritmos podrá con todo. Es la naturaleza estratégica, que subyace a muchos fenómenos de la política y la economía, lo que explica su impredecibilidad, no la falta de datos ni de algoritmos más eficientes. En el caso de los penales, y en lo que se refiere a predecir exactamente hacia dónde pateará un jugador, los datos previos solo parecen agregar ruido y confusión, como predice la tan clara como paradójica teoría de juegos.

En pleno auge de big data, machine learning e IA, seguimos sin haber mejorado nuestra habilidad predictiva de cosas aparentemente triviales. Porque munidos de los “consejos” de De Pablo o del amigo de mi padre, compradores y vendedores se encuentran en el mercado, apuntándose entre ellos como en el famoso meme de los hombres araña que pulula por las redes sociales.

Las nubes ignoran a los meteorólogos, pero la economía no ignora a los economistas

“Podrás engañar a algunas personas algunas veces, pero no a todas las personas todo el tiempo”, dicen que dijo Abraham Lincoln, frase que Bob Marley inmortalizó en “Get up, stand up”, el llamado a las armas que creó junto con Peter Tosh. Esta frase da en el centro de una de las razones detrás de por qué es muy difícil hacer predicciones confiables en economía, política, deportes o cualquier aspecto de la sociedad en que prevalezcan los comportamientos estratégicos.

Lamentablemente, la palabra “equilibrio” está “cargada” en la sociedad y admite muchas acepciones. Así, un equilibrista no se cae o, en el campo de la autoayuda, si alguien está “en equilibrio”, diremos que está tranquilo, estable, bien. En economía, y en las cuestiones sociales, el tema es un poco más complejo y alude a la relación que hay entre cómo la gente se comporta, cómo cree que funciona la sociedad y cómo todos creen en todo lo anterior.

Como sospecharán, no existe ninguna forma simple de referir a un concepto tan complejo que, para colmo, parece ser autorreferente: el equilibrio remite a sí mismo para existir. Por ejemplo, si yo tuviera una adecuada dosis de autoridad y capacidad de convicción, y les asegurase en este instante (y les explicase, argumentase o engañase) que exactamente dentro de cinco días habrá una crisis de proporciones bíblicas, y todo el Universo me creyese, es casi seguro que abandonarían la lectura de este fascinante libro para actuar en consecuencia: advertir a sus seres queridos, acopiar alimentos no perecederos, cargar el celular, intentar vender esto, comprar lo otro, lo cual provocaría la crisis de marras… ahora mismo. Con este cuento de las predicciones, los equilibrios y la credibilidad comunal hemos fabricado una hermosa contradicción: la creencia en la predicción y su timing terminan refutándola: el argumento (o capacidad de convicción) que es capaz de modificar acciones (dejar

todo y actuar) adelanta la crisis.

Sin entrar en ningún detalle técnico, la hipótesis de expectativas racionales se relaciona con estas ideas. Dudo que exista un principio social más complejo y discutido que esta hipótesis, y hacerlo con detalle excede por lejos los objetivos de este libro; de hecho, una gran idea sería dedicar un libro de divulgación científica a esta complejísima cuestión, créanme.

En un iluminador artículo, el microeconomista David Levine (una verdadera mente brillante) dice que el principio de expectativas racionales es algo así como el equivalente en la economía o la política del famoso principio de incertidumbre de Heisenberg en la física: es el principio que "regula" la relación entre el analista y aquello que es analizado. Lo hemos dicho antes: a las nubes les importa un rábano lo que diga la meteorología. Por el contrario, la opinión de un analista y la interacción con la sociedad puede tener impactos considerables en el propio funcionamiento de un mercado o una elección.

El principio de expectativas racionales postula que los agentes sociales (políticos, económicos, etc.) internalizan este doble rol de que quien analiza la realidad la altera con su análisis. Dicho de otra forma, que sus comportamientos son compatibles con este sistema de creencias múltiples y recíprocas.

En las cuestiones sociales, analizar la realidad no es simplemente pensarla para escribir un libro, dar una clase u opinar en un medio, sino una forma de posicionarse con respecto a lo que uno haría en relación con lo que el resto haría también. Patear y atajar un penal es un juego estratégico tanto como ganar o perder una elección, o vender o comprar una acción, y por eso los análisis importan; no son elementos ajenos al propio juego. Lanzar una pelota a un aro de básquet no es un juego estratégico, el resultado depende enteramente de lo que uno haga: al aro no le importa lo que uno haga o crea.

En relación con la capacidad de hacer predicciones con datos y algoritmos, la ausencia de estrategia no elimina el azar. Ciertamente, en lanzar una pelota a un aro existe el mismo tipo de azar que el analizado cuando hablamos del caso de la moneda en el

capítulo 1: el viento, la luz, el pulso del lanzador, alguna cuestión emocional, un perro que ladra son todos microfactores que influyen en el resultado más allá de la pericia del lanzador. Y todo esto constituye una dificultad en sí misma para la actividad predictiva.

Todos estos factores aleatorios se hallan presentes y de idéntica forma en los sucesos estratégicos. En la política, eventos como el covid-19, las acciones terroristas de Hamás o, en otra escala, un dicho inoportuno de un político o una tormenta que cae el día de un acto eleccionario pueden tener impacto en los resultados de una contienda política.

¿Atajar al azar?

Llegué a Río Cuarto, Argentina, a dictar una conferencia sobre la temática de este libro. Al mediodía, mis anfitriones habían organizado un almuerzo informal con los colegas del Departamento de Economía. Cuando comencé a comentar estas cuestiones relacionadas con el azar y la estrategia, Miguel Bosch, director de ese departamento, me contó una historia que ya comparto con ustedes. Miguel se reúne seguido con sus amigos de la infancia para jugar al póker. Nunca por plata, siempre por el más importante de los incentivos: el placer de ganarle a un amigo de verdad, porque con nadie se compite más fuerte que contra quien se admira y se quiere.

En una oportunidad, luego de varias partidas, uno de sus amigos le reveló a Miguel que "le había sacado la ficha" de que cuando tenía cartas buenas hacía un pequeño gesto que lo delataba. Me dijo Miguel: "A partir de ahí, decidí que cada vez que me vinieran cartas buenas, iba a tirar una moneda, e incorporar ese evento azaroso en mi decisión, para que no se revelara tan fácilmente mi estrategia". Y esto dispara una pregunta, que no le hice a Miguel: ¿en serio tirás una moneda o es como si lo hicieses? ¿Es realmente factible producir el azar?

El estudio de penales de Palacios Huerta, que comentamos antes, es importante por varias razones. Una es que se trata de la primera evidencia concreta de cuán relevantes son las estrategias mixtas, en las que hay explícita intervención del azar, como consecuencia de un equilibrio en que nadie quiere otorgar ninguna ventaja trivial a sus contrincantes. En juegos estrictamente competitivos, la aleatoriedad surge, espontánea, del intento de los jugadores de superarse unos a otros en astucia.

Para los fines de este libro, como dijimos, una importante consecuencia de este razonamiento es que limita el rol que cumplen los datos en los pronósticos.

Pero hay otra consecuencia, menos discutida pero igual de relevante en relación con las (in)habilidades de las personas para lidiar con el azar. Al pasar hemos dicho que, de acuerdo con los resultados de Palacios Huerta, arqueros y delanteros aleatorizan sus acciones para no dar ventajas en la competencia; atajan y patean como si lanzasen una moneda, como me dijo Miguel Bosch que hacía cuando jugaba al póker con sus amigos.

Por cierto, ningún delantero o arquero profesional entra al estadio con monedas, ni con otros dispositivos que le permita involucrar directamente al azar, a diferencia del árbitro, que sí lo hace para, por ejemplo, dirimir quién saca y de qué lado al comienzo del partido. Siempre según las demostraciones de Palacios Huerta, los jugadores involucran el azar en sus decisiones; entonces, lo hacen mentalmente o vaya a saber uno cómo. Lo cierto es que lo hacen, y que lo hacen bien. De por sí, este es un resultado sorprendente.

Vuelvo a mi pregunta: ¿es posible "provocar" el azar? Deberían sospechar que hay algo esencialmente contradictorio en provocar algo fortuito. Mi esposa parece opinar que sí. Lo explico. Mi casa es un auténtico museo de objetos que fueron apareciendo en la larguísima serie de viajes a los que nos llevaron nuestras profesiones. En particular, hay una mesita en el living ocupada por muchos de estos chirimbolos, testigos de una tensión familiar atávica: yo los ordeno casi maniáticamente (por filas, columnas, tamaño, etc.) y mi esposa se encarga de desordenarlos, para que parezca que fueron puestos al azar, o para que no se note que vive con un proto-serial killer como yo. Cada tanto, la sucesión orden-desorden es interrumpida por una diatriba como la que despliego a continuación, con la que intento convencerla de que es casi imposible hacer cosas al azar. Veamos por qué.

Cara de póker y “duro de aleatorizar”

No es tarea sencilla invocar al azar. De hecho, a las computadoras no les sale. ¿Cómo es eso? Bueno, creer o reventar. Cuando uno le pide a la compu que genere números al azar, lo que en realidad hace es disparar un algoritmo que en su estructura no tiene nada de fortuito (y que se parece a los sistemas caóticos que vimos en el capítulo 6), pero que se comporta como si lo fuese.

El matemático George Marsaglia es conocido por haber diseñado un conjunto de pruebas estadísticas para comprobar si realmente un algoritmo es capaz de producir números indistinguibles del azar. La batería de tests que diseñó se conoce con el oportuno y chacotón nombre de diehard (duro de matar), que me recuerda a cuando mi padre me contó que en la vieja Dirección General Impositiva había un empleado que tenía las citaciones de los evasores en una carpeta llamada “carne para la ganchera”.

Y si a las computadoras les es complicado generar números al azar, a los humanos les va bastante peor, como sutil y recurrentemente le recuerdo a mi esposa. Existe una copiosa evidencia (relatada en un hermoso artículo del notable matemático, divulgador y periodista argentino Adrián Paenza, cuya referencia aparece en el apéndice) de que aun a personas con abundante entrenamiento matemático-computacional les cuesta comportarse como un dado o una perinola.

Sin entrar en detalles, lo que deschava al humano es que tiende a confundir azar con erraticidad, ideas que discutimos con detalle en el capítulo 5. Es decir, si identificamos cara con 1 y ceca con 0, las sucesiones 1111111111 o 0000000000 son tan compatibles con el azar como 1011010010 o 0101101110. Digresión: para generar estas dos últimas series pensé en pedirle a mi esposa que lo intentase o hacerlo yo mismo, pero, a la luz de lo que quiero contar, se lo pedí a Roccatagliata, que es obediente y hacendoso con lo computacional.

Momento. ¿Cómo es que 0000000000 es tan compatible con el azar como 1011010010? Bueno, muchas veces la matemática (y la ciencia, si vamos al caso) se da de patadas con las intuiciones. Si, como ya aprendimos, "azar" implica tirar una moneda de modo que sacar cara o ceca tenga las mismas chances y que cada lanzamiento sea independiente de cualquier otro, las chances de que en dos lanzamientos salga 00, 01, 10 o 11 son todas iguales: 0,5 por 0,5, vale decir, 0,25. ¿Por? Porque las chances de que salga 0 o 1 en el primer tiro son del 0,5 y lo mismo para el segundo lanzamiento. Entonces, cualquier par de números tiene idénticas chances de salir en dos lanzamientos, y como hay cuatro resultados posibles, las chances son una en cuatro, ¼, vale decir, 0,25. Y así sucesivamente. Por supuesto, esto es difícil de tragar, y explica por qué el azar que gobierna a diez unos es el mismo que está detrás de 0101010101.

Por razones difíciles de entender (o no, la ciencia del comportamiento estudia con detalle estos procesos mentales), a la gente le parece que hay más azar detrás de 0101101110 que de 1111111111. Y de ahí que, si se pone a replicar el azar, tiende a romper las rachas a una tasa mucho más frecuente de lo que el mismísimo azar haría.

Como cuenta Paenza, si sin recurrir a un dado se pide a la gente que diga diez ceros o unos de modo que "parezcan al azar" (¡y pasen el test de Marsaglia!), la mayoría de las personas dice series de números que tienen pocas "rachas" (repeticiones de ceros y unos), ni hablar de que nadie diría diez unos ni diez ceros. Y eso es lo que discuto hace treinta y cuatro años con mi esposa: su "desorden" de los objetos es desorden, pero no azar.

El mensaje es que a la gente le cuesta mucho comportarse al azar, como los jugadores de fútbol o póker. Esto parece un detalle de nerds, para entretener a los alumnos y amigos, pero tiene una consecuencia severísima: si nos cuesta reproducir el azar, de igual modo nos cuesta reconocerlo, ni hablar de predecirlo. Y así es como mucha gente ve que en la ruleta ya salió siete veces el "rojo" y cree que, dado que salió muchas veces, debería deducirse que

ahora es más esperable que salga “negro”, para que todo vuelva a la normalidad;

el envión sugiere que va a seguir saliendo rojo, incluso si ya hemos razonado y explicado que, independientemente de lo que haya pasado antes, en el octavo tiro las chances de que salga rojo o negro no dejan de ser del 0,

Fuera de la ruleta y las moneditas, más de un emprendedor con chupines atribuye el éxito de sus inversiones a su visión y pericia, cuando tal vez solo tuvo suerte. Por el contrario, alguien con baja autoestima le asigna al azar la causa de sus penares, cuando quizás exista alguna razón sistemática, bajo su control o no, a la que culpar. De hecho, en las justificaciones detrás del éxito de la Argentina en el Mundial de Fútbol de Qatar se habla de la magia de Messi, del liderazgo de Scaloni y de la estirpe gaucha, pero no del azar, como si su presencia fuese en detrimento de lo anterior. Esta inhabilidad de internalizar el azar, de verlo, de evaluarlo, de darle el espacio que se merece (ni más, ni menos), es un tema delicadísimo que dificulta tanto la evaluación como la predicción de acciones, que van de lo personal a lo colectivo, del deporte a las políticas públicas.

¿Y qué entrevera a los arqueros y delanteros en esta disquisición sobre el azar y el derrotero de la vida? Que, según los resultados de Palacios Huerta sugieren, los jugadores profesionales –a diferencia de mi esposa, de los estudiantes de matemática y de casi todo el mundo– no parecen tener ningún problema en provocar el azar. Entonces, arqueros y delanteros aleatorizan sin problema alguno, sin apelar a moneditas ni –como sí hago yo– al ChatGPT.

Esto es sorprendente y difícil de explicar, y es una pregunta abierta. Puede que el contexto (el nerviosismo ante el penal, la energía de un partido difícil, etc.) los lleve a confiar en instintos o detalles ínfimos que provocan reacciones inmediatas e inconmensurables, que se parezcan al azar. Puede que la naturaleza hipercompetitiva haga que los jugadores profesionales no se dejen engañar con las

disquisiciones con las que nos distraemos los que opinamos desde la comodidad del teclado de una computadora. También, como me señaló el matemático Pablo Groisman cuando le conté esta historia, puede tener que ver con que arqueros y delanteros aleatorizan "de a un episodio por vez" y en un partido olvidan lo que hicieron en los anteriores, y quizás eso facilite la aleatorización misma, sin importar que el folklore del fútbol hable de papelitos, cábalas y otras costumbres que sugieren que los jugadores, más que olvidarse del pasado, se aferran a él como náufrago al salvavidas.

En el ámbito del póker, las estrategias de "bluf" o "faroleo" consisten en engañar al contrincante introduciendo un componente de azar. Un experimentado jugador me contó que, para no deschavarse, apela a la posición del segundero en el reloj, al color de alguna carta en la mesa o a si la cantidad de fichas son pares o impares para disparar o no estas estrategias, siempre con "cara de póker", obvio. No, no puedo revelar el nombre de mi informante, porque sigo fiel a mi conducta: ya les dije que este libro es científico pero no buchón.

De hecho, una interesantísima prueba sería pedirles a arqueros y delanteros profesionales que generaran sucesiones de unos y ceros, y ver si realmente pasan la prueba de "duro de matar" de Marsaglia, sobre todo teniendo en cuenta que las sucesiones de unos y ceros que surgen de verlos patear o atajar sí lo hacen, como muestra el estudio de Palacios Huerta.

Una vez más, en el marco de un libro de pronósticos como este, la capacidad de "abrazar el azar" (título del brillante libro de Pablo Groisman sobre el tema) es tal vez la habilidad más deseable que debe cultivar cualquiera que quiera plantarse frente al futuro. O, en solitario, delante del espejo del baño, a la hora de evaluar cómo le ha ido a uno en la vida. Porque a ningún ganador de la lotería lo invitarán a dar una charla TED para que cuente "cómo lo hizo". ¿O sí?

Volver del futuro

Este capítulo ahonda en la naturaleza predeciblemente impredecible de ciertos fenómenos de la vida cotidiana. Un primer sentido en el cual ciertos aspectos de la sociedad son impredecibles es cuando surgen como equilibrios de situaciones estratégicas.

Patear penales al azar no es una consecuencia de haber “tirado la toalla”, sino exactamente lo contrario. Es razonar que priorizar cualquier otra estrategia trivial necesita suponer que uno juega contra un idiota. O sea, la estrategia de involucrar al azar es una forma de evitar perder sistemáticamente.

Muchos eventos sociales, económicos, políticos y mundanos parecen disfrazarse para no ser descubiertos. Y por eso es muy difícil, si no imposible, pronosticar fenómenos en los deportes competitivos, las finanzas, la política, los negocios y muchos otros donde la estrategia mete su espantosa mano.

En los mercados, muchas veces los precios obedecen a un equilibrio en que las expectativas de los agentes (que compiten entre sí) están alineadas con la información disponible. Este equilibrio es dinámico y cambia con nueva información, lo que hace difícil predecir los movimientos futuros de los precios solo observando el pasado.

Para peor, discutimos que involucrar al azar (para evitar ser deschavado) no es una tarea trivial. No tenemos una forma obvia de

producirlo y, por consiguiente, nos cuesta verlo y predecirlo.

Tal vez uno de los yerros predictivos más comunes tiene que ver con darle al azar más o menos importancia de la que se merece.

10. Manejar mirando el espejo retrovisor: horóscopo, el diario de mañana y cisnes negros

En 1993 me mudé a los Estados Unidos a realizar mi doctorado. Decidí no llevar mi guitarra, porque creí que debía priorizar el estudio. Resulta que no soporté más de un mes sin poner mis dedos en las seis cuerdas, por lo que me lancé a buscar algún instrumento barato. Y así me hice de una vetusta Gibson Les Paul 1976: para la época, un catafalco, una "guitarra de viejo". Mis magros ingresos me llevaron a dejar pasar el modelo Custom, que en esos días salía unos pocos dólares más que la que yo me compré y todavía atesoro. Hoy, en 2024, el modelo Custom cuesta una fortuna y es un instrumento muy deseado por coleccionistas. Dejarla pasar fue la decisión financiera más estúpida que tomé en mi vida.

Obviamente, la evaluación de este episodio es ex post o, como se hizo costumbre decir en este libro, "con el diario de mañana". Este último capítulo versa sobre la tensión que hay entre predecir, describir y explicar, en particular eventos inexplicables o que ya ocurrieron. Veremos que racionalizar eventos del pasado puede ser una tarea útil para entender el futuro y también una fuente de errores predictivos, algunos muy groseros.

Pero antes de eso tenemos que hacer una parada técnica para discutir una fuente crucial de pronósticos de toda índole, que trasciende al tiempo, a los países, a las clases socioeconómicas.

Dios me libre de ser de Libra: la confiabilidad del horóscopo

Y en un libro de pronósticos tenemos que hablar del horóscopo, ¡no nos quedó otra! El de La Nación (que no me tutea) me dice: "Mercurio resalta la importancia de las negociaciones y el trabajo en equipo en asuntos financieros, así que busque asociaciones que beneficien a ambas partes". El de Clarín (que me tutea) me advierte: "Puedes experimentar estrés debido a soluciones que no llegan. Si esta situación se acentúa, deberás ser más precavido", y el de La Voz del Interior apunta: "Día para realmente tomarse el tiempo preciso como para poder organizarse internamente. Deja que los demás sean quienes tomen la iniciativa". Y como no encuentro una forma coherente de sacar una conclusión de toda esta información, le pido a Roccatagliata que opere su magia y que me resuma todo lo anterior en una frase. Obediente, dice: "Negocia en equipo y evita estrés decisivo" (sic). O sea, el consejo que todo el mundo le da a todo el mundo en toda circunstancia todos los días, como el "suprimí los postres" a los que empiezan a hacer dieta y el –por qué no– "buscate un trabajo honesto" de Pappo.

La pregunta obvia es cómo se explica la enorme popularidad del horóscopo y la astrología, que trascienden a la educación, la cultura y el estatus socioeconómico. Chequeo la página web del Diario Popular, y el ícono del horóscopo aparece en un lugar destacado, junto al de "lotería y turf". Lo mismo ocurre en la app de La Nación, el histórico y aristocrático diario argentino, pero en la paqueta sección lifestyle, acorde con su alcurnia (como dice la canción de Sui Generis). Milanga o "fritez de loncha bovina vulnerada en colchón de huevo y astillas de pan", milanesa al fin: el horóscopo está y estuvo en todas partes. Y por algo será.

Hay dos conjeturas detrás de la explicación de la popularidad del horóscopo. Una es que la astrología tiene un récord relevante de predicciones correctas, y de ahí su aceptación masiva. La otra es "porque no se dieron cuenta de que no funciona". Está en el ADN

del científico social (a la larga, soy un científico social) rechazar este tipo de explicaciones. Aceptar cierto grado de inocencia general y persistente es como decir “porque sí” o, en el mejor de los casos, apelar a un eufemismo para barrer el problema debajo de la alfombra: si mucha gente no se da cuenta de algo, eso en sí mismo es un patrón social relevante. Entonces, hagamos un pequeño recorrido por las razones detrás de la relevancia de la astrología y el horóscopo.

La primera explicación indaga “si el horóscopo y la astrología funcionan”. Rápidamente podríamos caer en una “madriguera de conejo” (por la expresión rabbit hole, en inglés) semántica, y pasar estas páginas casi finales del libro discutiendo qué significan “horóscopo”, “astrología” y, sobre todo, “funciona”, y quedar atrapados en ella.

Históricamente, los disensos acerca de estos conceptos gobernaron los debates alrededor de esta cuestión. Los estudios llevados a cabo por científicos eran criticados por la comunidad de astrólogos sobre la base de que lo que estaban evaluando no era exactamente la astrología; y los implementados por los astrólogos eran desestimados por los científicos por no respetar las pautas metodológicas de las evaluaciones de la ciencia. También eran comunes las argumentaciones (o falacias) ad hominem que, más que en discutir datos, definiciones o métodos, se enfrascaban en denostar a quienes las sostenían por cuestiones meramente personales, tanto de parte de los astrólogos como de científicos. Piñas van, piñas vienen.

Esta disquisición tuvo un quiebre fundamental en el año 1985, con la publicación de un revolucionario (y ampliamente citado) estudio del físico estadounidense Shawn Carlson, cuando todavía era un joven estudiante de licenciatura. Carlson (¡de entonces 19 años!) se propuso la quimérica tarea de hacer que científicos y astrólogos concordaran en varias dimensiones: qué es la astrología, cuál es su tesis fundamental, cómo debería evaluarse y qué se entiende por una aseveración tal como “la astrología funciona” (o no). A tal efecto, logró que la comunidad científica (representada por los físicos y estadísticos de la prestigiosa Universidad de California, en Berkeley) interactuasen con los astrólogos, agrupados en el Consejo

Nacional de Investigación Geocósmica de los Estados Unidos.

Ambas partes acordaron en que la tesis fundamental de la astrología natal es la siguiente (tomada del artículo de Shawn):

Las posiciones de los planetas (todos los planetas, el Sol y la Luna, además de otros objetos definidos por los astrólogos) en el momento del nacimiento se pueden usar para determinar las características generales de la personalidad del sujeto y las tendencias en su temperamento y comportamiento, y para indicar los principales problemas que es probable que este enfrente,

es decir, si las aseveraciones con las que comenzamos esta sección tienen algún sentido real, si hacen predicciones confiables o si describen alguna característica objetiva de las personas.

La herramienta convenida fue un experimento “doble ciego”, en el que las partes involucradas en el estudio no se conocen entre sí. Shawn propuso dos experimentos. En el primero, se eligió a un grupo de personas al azar, utilizando un puntilloso mecanismo como los que se usan para experimentos en medicina o psicología, y se les pidió que diesen la mínima información necesaria para que 28 astrólogos confeccionasen e interpretaran sus cartas natales. Esas cartas fueron confeccionadas con la ayuda de una computadora (todo un avance, para la época) y la interpretación e informe final de las cartas fueron realizados por los astrólogos. Refuerzo un poco la explicación. “Doble ciego” quiere decir que ni los astrólogos sabían quiénes eran los voluntarios, ni los voluntarios quiénes eran los astrólogos, una práctica habitual en experimentos científicos.

A cada voluntario se le mostraron tres cartas astrales y sus informes: uno de ellos, el que le correspondía, y otros dos, elegidos al azar de entre los que se habían elaborado. A cada sujeto se le pidió que eligiera cuál creía que le correspondía. Si la astrología no tuviese ningún impacto predictivo, los participantes elegirían correctamente su propia carta natal solo un tercio (poco más de un 33%) de las veces, y si la astrología funcionase a la perfección, lo

harían bien en el 100% de los casos. Los astrólogos y los científicos acordaron de antemano que estarían dispuestos a creer que la astrología funcionaba si la tasa de acierto era, como mínimo, un prudente 50%.

El proceso de recolección de datos e implementación del experimento fue un auténtico calvario, tanto por los celos de los bandos involucrados como por una sucesión de extraños sucesos, incluido el caso de un "voluntario" que resultó ser un astrólogo encubierto, y otros dos que en el transcurso de la prueba (que duró varias semanas, en época preinternet y con idas y vueltas de correo) se convencieron de que "la astrología era un trabajo demoníaco y rehusaron participar en lo que ellos creían que era "un test experimental del diablo" (sic). El artículo de Carlson (publicado en la prestigiosísima Nature) describe con minuciosidad todo el proceso de recolección y análisis de datos.

Los resultados son impresionantes: solo el 33% de los voluntarios pudo elegir correctamente su carta astral, muy por debajo del 50% esperado por los astrólogos, y casi igual al tercio predicho por los científicos.

Acaso alguien objete que las personas no tienen la habilidad de reconocer sus propios rasgos de personalidad, lo cual puede introducir en el experimento una considerable dosis de azar, cuando no sesgos. A fin de afrontar ese reparo, Carlson, los científicos y los astrólogos acordaron implementar un segundo experimento, en el que a cada voluntario se le administró el entonces famoso California Personality Inventory (CPI, Inventario de Personalidad de California), una medida estandarizada de rasgos de personalidad diseñada por psicólogos. Ahora los roles se invirtieron. A cada astrólogo le dieron la carta astral de un voluntario, el resultado de su CPI y dos más tomados al azar. A cada astrólogo se le pidió que identificara el correspondiente a la carta astral recibida, siempre en el marco del diseño doble ciego. Los resultados fueron virtualmente idénticos a los del experimento anterior: los astrólogos solo acertaron el 34% de los casos, casi exactamente lo que predijeron los científicos escépticos.

La categórica conclusión de Carlson es que

pese a haber trabajado con los mejores astrólogos del país […] y haber incorporado todas sus sugerencias, […] su predicción de las conexiones entre las posiciones de los planetas y otros objetos astronómicos al momento del nacimiento y las personalidades de los sujetos experimentales no existe. El experimento claramente refuta la hipótesis astrológica.

Como es de esperar, las críticas de parte de la comunidad astrológica no tardaron en llegar y cuestionaron desde la forma en la que operó el experimento hasta aspectos personales de Carlson, además de atacar a la psicología, poner en entredicho la relevancia de los tests implementados, la ubicuidad de un análisis cuantitativo cuando uno cualitativo era más apropiado, la idoneidad y honestidad de Nature y su editor para actuar como referees de esta cuestión, incluido el argumento de que la astrología no es una ciencia sino un arte y que, como tal, la evaluación de Carlson resultaba inadecuada. En su revisión de críticas al estudio de Carlson, el astrólogo Robert Currey concluyó: "Los astrólogos pueden citar legítimamente el estudio de Carlson publicado en la revista científica más prestigiosa, Nature, como evidencia adicional que respalda la tesis de que si bien la práctica de la astrología es sobre todo un arte, el campo de la astrología natal parece tener cada vez más una base científica", además de criticar enfáticamente a Wikipedia por dar demasiada relevancia a las conclusiones de Carlson.

Como les adelanté, se pueden esgrimir varias razones acerca de por qué la gente cree en el horóscopo; les advertí que las explicaciones "porque no se dan cuenta de que no funciona" son inaceptables para cualquier científico social serio. Los sesgos de confirmación, de los que ya hablamos, son una de las racionalizaciones del horóscopo. La vaguedad del horóscopo y la propensión a creer en él hace que prioricemos la información que confirma la hipótesis de que el horóscopo funciona. La psicología y, en general, las ciencias del comportamiento han dedicado considerable espacio a explicar por qué tenemos este tipo de reacciones "raras".

Para explicar la creencia en el horóscopo también se cita recurrentemente el “efecto Barnum” (o Forer), que se refiere a la tendencia a personalizar aseveraciones generales de descripciones genéricas de la personalidad. En 1948, el psicólogo Bertram Forer administró a sus estudiantes un test de personalidad. A posteriori, cada alumno recibió una descripción de su personalidad individual, sobre la base de los resultados del test. Cada descripción era, en realidad, una serie de declaraciones generales y ambiguas que podrían aplicarse casi a cualquier persona (¿como las del horóscopo y las de las galletas de la fortuna de los restaurantes chinos?) A pesar de la falta de especificidad, en su mayoría los estudiantes consideraron que la descripción era precisa y aplicable a ellos mismos.

* * *

Hace varias horas que trabajo en esta sección del libro, que resume lecturas y entrevistas que me tomaron varios días de indagación, incluida una desopilante encuesta en redes sociales en la que el 86,4% de los encuestados me dijo que no creía en la astrología o el horóscopo. Hoy me largué a escribir a eso de las 8 de la mañana; ya son exactamente las 3 de la tarde y no me doy cuenta de cómo cerrar este apartado, lo que me genera alguna angustia. Y ahora que me fijo, el horóscopo de Clarín tenía razón cuando hoy temprano me advirtió: “Puedes experimentar estrés debido a soluciones que no llegan. Si esta situación se acentúa, deberás ser más precavido”. Y veo que La Voz también acertó cuando acudí a mis seguidores en Twitter para dirimir esta cuestión. Entonces, me dispongo a seguir los consejos astrales. Antes de seguir adelante, me voy a hacer un té y a apostar por la precaución, como recomendó Clarín. Ahí vuelvo.

Kafka llegó hace rato: pronosticar el pasado

Un error muy usual en el ámbito de las predicciones es creer que uno pudo haber predicho algo por el mero hecho de haber encontrado una racionalización ex post. Es la típica predicción… con el diario de mañana, que no sirve como predicción y, en general, tampoco como explicación. Este error se magnifica si refiere a eventos esencialmente impredecibles, como los penales y los "paseos al azar" del capítulo anterior, o los "cisnes negros", que nos ocuparán más adelante en este capítulo.

Así, muchos dicen haber anticipado que Jorge Bergoglio iba a ser el nuevo papa, aun cuando un día antes del 13 de marzo de 2013, cuando fue designado, casi nadie lo tenía en mente. A modo de ejemplo, el 12 de marzo de ese año, el diario La Nación tituló: "El italiano Scola es el favorito de los apostadores para ser elegido papa" y debajo, en letra microscópica: "Los argentinos Sandri y Bergoglio, con pocas expectativas". Después aparecieron los expertos, los analistas, los "visionarios", a explicarlo todo.

Tal vez una de las mejores reflexiones sobre las limitaciones de armar historias sobre el pasado y predecir con el diario de mañana se halle en "Kafka y sus precursores", un inspirado ensayo de Jorge Luis Borges publicado en 1951 (y en 1952 integrado, con retoques, a Otras inquisiciones). Borges comienza su texto refiriéndose a Kafka: "A poco de frecuentarlo, creí reconocer su voz, o sus hábitos, en textos de diversas literaturas y de diversas épocas". Tras lo cual pasa revista a varios autores que, en cierto modo, califican como predecesores de Kafka. Nada demasiado llamativo, hasta ahora.

Pero luego don Jorge Luis frota su prodigiosa lámpara para cuestionar hasta qué punto la existencia de predecesores de Kafka no depende estrictamente de la existencia y la lectura de… Kafka. ¿Hay tal cosa como "los predecesores de Kafka" o es la literatura del notable autor checo la que, estudiada ex post, permite identificar a sus antecesores? Dice Borges:

El poema Fears and Scruples de Robert Browning profetiza la obra de Kafka, pero nuestra lectura de Kafka afina y desvía sensiblemente nuestra lectura del poema. […] El hecho es que cada escritor crea a sus precursores. Su labor modifica nuestra concepción del pasado, como ha de modificar el futuro.

El argumento de Borges refiere a una situación que concierne a muchos fenómenos, tanto de la ciencia como de la vida cotidiana: ¿hasta qué punto valen las explicaciones "con el diario del lunes"? En el marco de ese ensayo, en su momento, ¿habría sido posible detectar la naturaleza predecesora de Zenón, Han Yu, Kierkegaard, Browning y los otros que Borges menciona como precursores del autor de La metamorfosis? El ensayo mismo sugiere que no, porque la esencia de ser precursor depende de la existencia, obviamente posterior, del propio Kafka, cuya obra y lectura crean a sus precursores. "En el socialismo nunca sabes el pasado que te espera", escribió Leonardo Padura, el notable escritor cubano, en relación con cierta tendencia a fabricar historias compatibles con un presente, haciendo correr el tiempo hacia atrás.

A la larga, y por alusión, el ensayo de Borges da de lleno en la complejidad (y, a veces, futilidad) de desandar el sendero que siguen ciertos procesos disruptivos, lo que nos enfrenta a una situación paradójica: ciertos fenómenos (los textos icónicos de Kafka, por ejemplo) son impredecibles ex ante y, a la vez, racionalizables ex post. Antes de Kafka había autores y textos, pero en algún momento apareció Kafka, no como una continuación trivial de sus precursores sino como una pieza clave e inadivinable, en grado tal que, paradójicamente, ciertos textos lo anteceden y casi lo vaticinan, pero es la literatura de Kafka la que permite identificarlos como tales y no al revés.

Inquirido acerca de cuál es el mejor libro de cálculo, el notable matemático y pedagogo francés Serge Lang dijo: "El cuarto que uno lee", reflejando que el aprendizaje de una disciplina compleja, como la matemática, está plagado de discontinuidades, no linealidades y elementos fortuitos que reclaman una borgeana práctica circular en

la que un tema es visitado y revisitado hasta que, en algún momento, cobra sentido (cuando uno leyó el cuarto libro, siguiendo la idea de Lang) y pasa de imposible a trivial. Todos, en la secundaria o en la facultad, hemos lidiado durante horas o días con un problema aparentemente imposible que, de pronto, casi como por arte de magia, cede a nuestros intentos en medida tal que luego nos resulta difícil entender por qué demandaba semejantes esfuerzos.

Es esa discontinuidad en el proceso de aprendizaje lo que complica hacer "ingeniería reversa" del camino obtuso y discontinuo que nos llevó a entender un fenómeno complejo. Con relación a Browning (uno de los precursores de Kafka), Borges dice: "Browning no lo leía como nosotros ahora lo leemos"; es la existencia (posterior) de Kafka lo que altera la lectura de Browning. Ajenos a esta importante apreciación, un error (muchas veces involuntario) de muchos docentes es recomendarles a colegas y alumnos "el cuarto libro que leyeron" porque, a la luz de un conocimiento posterior (haber aprendido), juzgan que es el mejor, cuando ese estatus depende de haber leído antes los otros tres y, fundamentalmente, de haber comprendido un tema. La impredecibilidad de Kafka solo sobre la base de sus precursores se parece a la dificilísima tarea de comprender cómo era la cabeza de uno antes de entender un problema complejo, cómo era un mundo sin Kafka, aun cuando sus predecesores ya andaban pululando por ahí; cómo se explica un Kafka, cómo se crea un Kafka, cómo se predice un Kafka (un Messi, un Piazzolla, un Borges).

En relación con la tarea de pronosticar, hay dos corolarios interesantes que se desprenden de la aguda reflexión de Borges. Primero, que ni el propio Kafka es necesariamente un precursor de sí mismo, como sugiere Borges al finalizar su texto: "El primer Kafka de Betrachung es menos precursor del Kafka de los mitos sombríos y de las instituciones atroces que Browning" (de paso, digamos que Betrachung es el libro Contemplación). Segundo, que el razonamiento sugiere que, en un sentido causal y no temporal, Kafka es un precursor de sus precursores.

Este es un problema recurrente en la tarea de pronosticar: el análisis de eventos presentes da señales claras de hechos pasados que

podrían haberse usado para el pronóstico, pero que, por la índole circular que nos sugiere el inspirado ensayo de Borges, son de uso limitado o inconducente.

Todos conocemos la historia de Van Gogh: un genio loco y traumado de vida terrible, cuya obra se apreció solo luego de su muerte. Entonces, con el diario de hoy, el negocio parece consistir en tomar la máquina del tiempo, volver a la Europa de la segunda mitad del siglo XIX, comprarle a Van Gogh sus cuadros por unas pocas monedas y revenderlos ahora arriba de 100 millones de dólares, como dicen que cotizan sus obras más famosas.

El problema es que, sin "el diario de mañana", el mandato no es comprar cuadros de Van Gogh, sino de cuanto loco pintaba cuadros en ese entonces, y ahí no resulta claro si realmente es negocio. Más simple aún, sabiendo que salió el 17 en la ruleta, de volver el tiempo atrás, el dato es "jugar a la ruleta": si uno les pone fichas a todos los números, seguro va a ganar, a costa de perder todas las fichas que les puso a los números que no salieron, que es lo mismo que no ganar nada.

Estas situaciones recuerdan a la ya mencionada frase del Indio Solari: "El futuro llegó hace rato". Sucede que la complejidad del nexo causal entre el pasado y el presente hace que sea difícil, cuando no imposible, reconocer esas señales que anteceden a un fenómeno relevante, como a los precursores de Kafka antes de Kafka.

Haciendo llover con paraguas

Peor aún, en muchas circunstancias es posible “predecir un Kafka” pero no explicarlo. Este es otro error predictivo en el que, lamentablemente, caen muchos (sobre)entusiastas de big data y la IA: creer que porque pueden pronosticar razonablemente bien un resultado pueden explicarlo, o simplemente pensar que están más cerca de hallar una explicación.

A fin de entender esta cuestión, exploremos la siguiente historia. Roberto es guardia de seguridad de un inmenso edificio público. Su tarea consiste en pararse junto a los ascensores, ubicados en el centro del edificio, desde donde no se divisa la calle, y velar por el orden y la tranquilidad. Él hace su trabajo con esmero, y espera que sus jefes lo noten y se dignen a trasladarlo del turno nocturno al diurno. Uno de sus entretenimientos es predecir el estado del tiempo viendo cómo ingresan los trabajadores y visitantes al edificio. Con una precisión de la que se siente orgulloso, Roberto sabe que si a eso de las 8 de la mañana, cuando está por terminar su turno, más de la mitad de la gente entra al edificio con paraguas, es porque habrá un largo día de lluvia y no será cuestión de un chaparrón aislado.

A cualquiera que haya pasado por el sistema escolar le han contado cómo funciona esto de la lluvia: que el agua se evapora, que sube y forma nubes, que se vuelven lo suficientemente pesadas y caen en forma de lluvia, y que Armando Manzanero escribe un bolero. En el cuento de las causas y los mecanismos de la lluvia no aparecen los paraguas, aunque, desde la perspectiva de Roberto, los paraguas anteceden a la lluvia. Dicho de otra manera, es la lluvia lo que lleva a las personas a usar paraguas, pero resulta que ver gente con paraguas permite predecir que va a llover, como hace Roberto.

Esta discusión apunta a que una cosa es que un fenómeno cause otro, y otra que lo anteceda. Acá la lluvia causa los paraguas y no al revés, si bien los paraguas la anteceden y facilitan el pronóstico,

como se dio cuenta Roberto.

En 1969 el economista Clive Granger propuso un mecanismo estadístico para detectar si un fenómeno precede temporalmente a otro, y la comunidad científica no tuvo mejor idea qué bautizarlo "test de causalidad de Granger". De hecho, si uno tuviese datos de la cantidad de gente que usa paraguas y de cuánto llovió, para un período largo, encontraría que el test de Granger valida esta conjetura. Pero la gente es mala y simplifica demasiado y, aun en este contexto obvio, pretende ahorrar tinta al decir (y muchas veces, creer) que estos resultados implican que "los paraguas causan lluvia". Para complicar más la cosa, más de un inescrupuloso u oportunista compra (o vende) la idea de que, "sobre la base de los resultados de un estudio científico", el Estado debería prohibir el uso de paraguas para que pare de llover o repartir paraguas en el desierto de Atacama para provocar chaparrones.

El propio Granger se ocupó de aclarar que hablar de "causalidad de Granger" es peligroso e inconducente, toda vez que su procedimiento consiste simplemente en un mecanismo para encontrar precedencia temporal en los datos y no causalidad.

En efecto, el test de Granger y sus herramientas asociadas son útiles para detectar precedencias temporales que puedan utilizarse para realizar pronósticos relevantes. La construcción de indicadores líderes es un claro ejemplo. Este tipo de indicadores ocupa un lugar importante en el quehacer económico. Sobre la base de métodos de series temporales como los analizados en este capítulo, estos índices intentan trackear el derrotero agregado de una economía. Por lo general, se construyen índices de tres tipos: líderes, coincidentes y rezagados. Los líderes son los más relevantes, ya que se adelantan a los ciclos y se usan para predecir comienzos de crisis o períodos de recuperación. Algunas variables clave que conforman este índice son datos de horas trabajadas por semana, encuestas de expectativas, índices de habilitación de negocios o construcción, etc. Los coincidentes son indicadores que, en cierto sentido, monitorean la evolución macroeconómica y se basan en variables como valores de empleo, valores de producción, entre otros. Por último, los rezagados se usan para confirmar o chequear tendencias.

En la Argentina, la Universidad Di Tella elabora un índice líder,

sobre la base de las siguientes variables: el índice de la Bolsa de Buenos Aires, el índice Merval, superficie por construir, despacho de cemento al mercado interno, índice de confianza del consumidor, concursos y quiebras, escrituras de compraventa de Buenos Aires, metros cuadrados construidos, índice de la industria metalmecánica y de utilización de capacidad instalada de bienes intermedios.

Una vez más, es importante aclarar que los indicadores líderes solo pretenden anticiparse a la realidad, no necesariamente explicarla ni usarla para pensar escenarios contrafácticos y actuar en consecuencia. Es decir, si bien el índice líder puede predecir el inicio de una recesión, no necesariamente es una herramienta útil para actuar a partir de ese dato, porque no surge de una explicación causal de cómo funciona la economía, sino de algunas ideas simples que permiten predecirla. Más aún, la literatura científica de índices líderes ha avanzado considerablemente en los últimos cien años, con mejoras notables en la capacidad predictiva, pero la naturaleza cíclica de las economías sigue teniendo efectos negativos indeseables. O sea, se ha avanzado mucho en predecir ciclos económicos, pero menos en explicarlos.

En la construcción de índices líderes y sistemas de monitoreo social, el espacio para big data y la IA es enorme. El proyecto Economic Tracker liderado por el economista Raj Chetty (firme candidato al Nobel) muestra cómo es posible monitorear la economía en tiempo prácticamente real y con una “granularidad” asombrosa (casi a nivel de comuna, o condado, en los Estados Unidos), usando información de big data, como las transacciones online o los datos inmediatos de movilidad (transporte, correos, encomiendas), que han sido utilizados para medir el derrotero de fenómenos urgentes, como los provocados por el covid-19. Los trabajos del argentino Alberto Cavallo muestran que datos “chupados” de internet pueden trackear, y a veces pronosticar, la inflación en frecuencia altísima (diaria, por ejemplo), tarea imposible con el sistema de medición actual usado por el Indec argentino, que con encuestas tradicionales solo puede monitorear la inflación de mes a mes.

Y así como coexisten los usos con los abusos de la tecnología, a veces la comunidad científica recurre inteligentemente a esta

mezcolanza intelectual con fines pedagógicos y también para divertirse un poco. En un desopilante paper, Walter Thurman y Mark Fisher, de la Universidad Estatal de Carolina del Norte, apelan al test de Granger para estudiar “la” pregunta causal-temporal por excelencia: ¿qué vino primero, el huevo o la gallina? Los autores recopilaron una larga serie de producción de huevos y gallinas, y sometieron los datos al procedimiento creado por Clive Granger. El estudio encuentra que son los huevos los que anteceden a las gallinas, dando por cerrado un eterno debate histórico.

El trabajo de Thurman y Fisher concluye con el desafío de usar el test de Granger para dilucidar otras cuestiones causales atávicas, como si es cierto que “quien ríe último ríe mejor”. Este libro sensible no puede quedar impávido ante semejante provocación. A tal fin, procedimos a analizar las búsquedas semanales de “risa” y “bienestar” en Google, para la Argentina, durante 2019 y 2024. Aplicado a estos datos, el test de Granger es concluyente: es el bienestar lo que provoca risa. Así, no es que el que ríe ultimo ríe mejor, sino que el que ríe mejor, ríe ultimo. Siguiente pregunta, por favor.

Un cisne negro de 23.388 plumas

El papa Francisco, Javier Milei, Trump, los Beatles, el covid-19, el Brexit, los atentados a las Torres Gemelas, la canción “Despacito”, la caída de Lehman Brothers son todos eventos “raros” y que tuvieron un impacto profundo en la sociedad. Como hemos analizado, el conspirativo de siempre lo previó todo; pero ya hemos dicho que estas “predicciones” (con el diario de mañana) son artilugios discursivos (para impresionar al grupo de WhatsApp de los vecinos) antes que conocimiento útil.

Claramente, algún sector de la comunidad vio la inminencia del covid-19, y tal vez con explicaciones válidas, lo que plantea una delicadísima cuestión en el marco de las predicciones: las predicciones que importan son las que convencen. Si ahora mismo yo soy capaz de ver la irrupción inminente de un fenómeno (un cataclismo, un accidente, la emergencia de un tirano o la letal llegada de la canción sucesora de “Despacito”) y no tengo forma de convencer a alguna parte relevante de la comunidad de mi hallazgo, no es obvio que tenga una predicción. Una frase clave de la estadística es que la técnica que nadie usa no sirve para nada. Algo parecido aplica a las predicciones: el pronóstico que no convence es como si no hubiese existido nunca.

Por consiguiente, la habilidad de realizar predicciones útiles tiene un enorme componente comunicacional. Desde ya, esta capacidad de convicción es comunal. Quizás la inminencia de un desastre natural primero tenga que ser comunicada en términos técnicos para convencer a los especialistas que, en lo sucesivo, deberían provocar una “cadena de convicción” que se derrame al resto de la sociedad, comenzando por los científicos y terminando por los periodistas, los divulgadores y los vecinos del barrio, todos contribuyen.

Pero lo cierto es que, aun en tiempos de big data, IA y comunicaciones veloces, ciertos eventos raros e importantes, como

los descriptos más arriba, ocurren con una frecuencia superior a lo que la mayoría espera.

El futurólogo-estadístico-trader libanés Nassim Taleb ha hecho una carrera de argumentar acerca de la relevancia de estos eventos raros, a los que bautizó como cisnes negros. La verborragia y la vehemencia de Taleb hacen que algunas de sus interesantes ideas terminen escondidas en ese torrente oratorio. Lo de cisnes negros viene de una famosa historia. Hasta el siglo XVII se pensaba que todos los cisnes eran blancos; pero cuando en Australia se descubrieron cisnes negros, el conocimiento existente quedó en entredicho.

Una forma de ver lo que dice Taleb es que, más que "la excepción que justifica (o pone a prueba) la regla", estos eventos exóticos son ellos mismos parte de una regla, más compleja de lo que la sociedad sospecha; o bien que relegarlos a la categoría de "excepción" hace que nos perdamos entender la naturaleza disruptiva de muchos fenómenos cruciales de la sociedad.

Es muy difícil definir qué significa "raro", porque cualquier definición precisa termina especificando qué es "normal" En ese sentido, Luca Prodan, el mítico líder de Sumo, cantaba "yo estoy al derecho, dado vuelta estás vos". Ya hemos enfrentado estas cuestiones aparentemente paradójicas: hay algo un tanto contradictorio en definir que algo es extraño, porque si podemos definirlo con detalle, quizás tan raro no era.

Una forma de afrontar esta delicada cuestión es hacerlo en términos relativos, es decir, raro es algo que ocurriría con bajísima probabilidad y que está lejos del "centro" en el que ocurre el resto de los fenómenos. Así, raro era Gheorghe Mureșan, un rumano que jugó en la NBA en los noventa y que medía 2,31 metros de altura.

Antes, una pequeña aclaración. Supongamos que la altura exacta de Mureșan es de 2,318765 metros, ahora medida con una precisión de seis decimales. Comparémosla con mi altura, de 1,652567 metros. Desde el punto de vista de la altura exacta, una persona de altura idéntica a la de Mureșan es tan rara como una de mi altura. La rareza que estamos discutiendo no se refiere a esta cuestión, sino al hecho de que personas con alturas parecidas a la mía son mucho

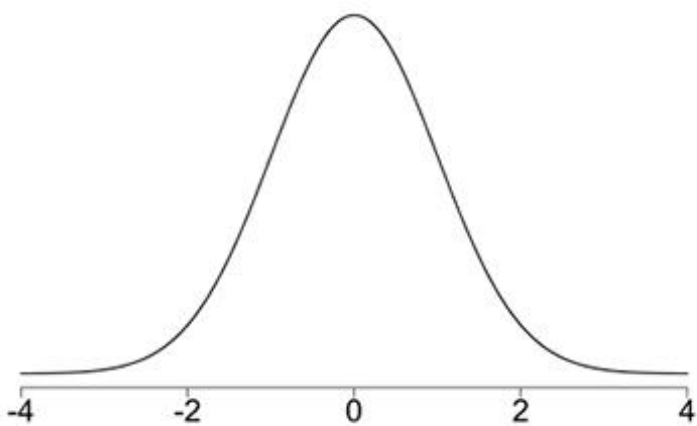
más abundantes que las similares a la de Mureșan. La rareza de la altura de Mureșan tiene que ver con que está muy lejos del promedio. Raro es Mureșan, no yo, ¡tomá mate!

Lo que discute Taleb se relaciona con cómo nos damos cuenta de que algo está lejos o cerca y qué significa "raro". Cuando yo tenía 6 años, mi idea de "cerca" era todo aquello que estaba en la cuadra en la que vivía. Desde esta perspectiva, Mar del Plata (a 404 km de Buenos Aires) estaba tan "lejos" como Shanghái (a 19.347 km). O sea, en definir "lejos-cerca" hay alguna dosis de discrecionalidad, que peligrosamente se transforma en arbitrariedad.

Una forma de saldar esta cuestión es apelar a la siguiente idea: "raro" es aquello que, estando lejos del centro, ocurre con bajas chances. Lo cual nos pone a considerar cómo se calculan las chances de las cosas, problema con el que hemos lidiado en el capítulo 1, no sin dificultades.

En relación con la altura de las personas, la forma más popular de resolver esta cuestión es acudir a la mítica distribución normal, de campana o gaussiana, que también mencionamos antes y que aquí muestra la figura 10.1:

Figura 10.1



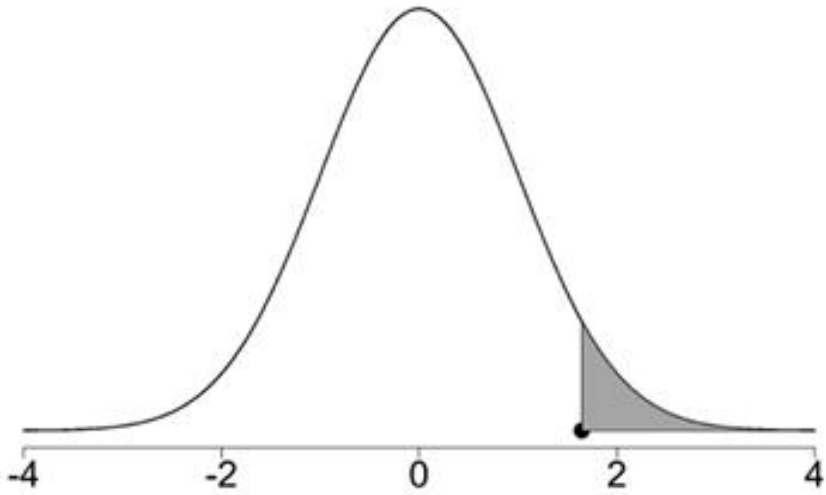
¿Cómo leer esta curva? Retomemos al caso de las alturas: piensen en todas las personas que conocen, tanto por contacto directo como por la tele, internet, los diarios, etc. Una recurrencia es que hay muy pocas personas extremadamente altas (como Mureșan) o bajas (como el comediante Danny DeVito, de tan solo 1,47 m de altura). Por otra parte, hay muchísimas personas de altura similar a la del promedio (1,73 m para los varones y 1,62 m para las mujeres, dice Roccatagliata). Dos cosas más. Y se va la primera. Es igual de extraño que aparezcan personas muy altas o muy bajas. Llega la segunda, fundamental. Si empezamos en el promedio, la cantidad de personas con altura parecida al promedio mismo cae en un comienzo de forma más o menos lenta y luego rápidamente.

Si la campana de Gauss de la figura 10.1 pasa a representar a las alturas de nuestra especie, el centro de la campana se corresponde con la altura promedio, donde uno debería esperar que se concentre la mayor cantidad de personas. Que las colas de la campana caigan de forma simétrica y lenta al principio y rápido después (y que justifica la forma de "campana") tiene que ver con lo que describimos en el párrafo anterior acerca de la manera como desaparecen las personas de altura parecidas al promedio.

Este tipo de curvas se llaman curvas de densidad en probabilidad y tienen una propiedad muy interesante. El área bajo toda la curva es 1 y el área para cualquier intervalo es la probabilidad de que ocurran valores en ese intervalo.

Entonces, si decimos que raro es que alguien sea muy alto y que aparezca con baja chance –por ejemplo, un 5%–, lo que hacemos es buscar el valor de la recta que deja a la derecha al 5% de todas las personas. Para ponerlo en evidencia, aquí está la figura 10.2.

Figura 10.2



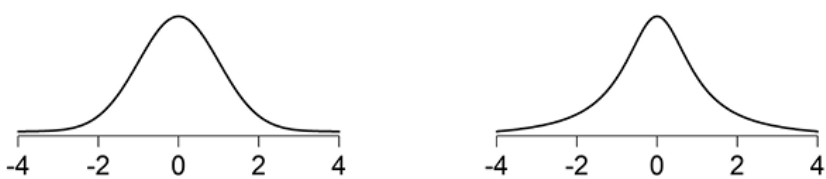
El punto marcado en el eje horizontal deja a la derecha al 5% de las personas; ahí están los altos, los raros. Para los Estados Unidos, este número da 1,85 metros. Solo el 5% de los varones adultos miden más que ese 1,85, son “altos”; el resto, no. En esta métrica, Mureșan no es raro, sino rarísimo.

¿Cuán normal es la distribución normal?

Con enorme acierto, Taleb pone de relieve una cuestión: más allá de su omnipresencia científica y cultural, muchos eventos de la sociedad no se rigen por la hermosa curva de campana de Gauss. Lo que dispara la obvia pregunta: ¿cuán normal es la curva normal?

En el siguiente dibujo –la figura 10.3– les muestro dos curvas de densidad. Una es la normal y la otra… es un misterio. ¿Se animan a decir cuál es la gaussiana-normal y cuál la misteriosa?

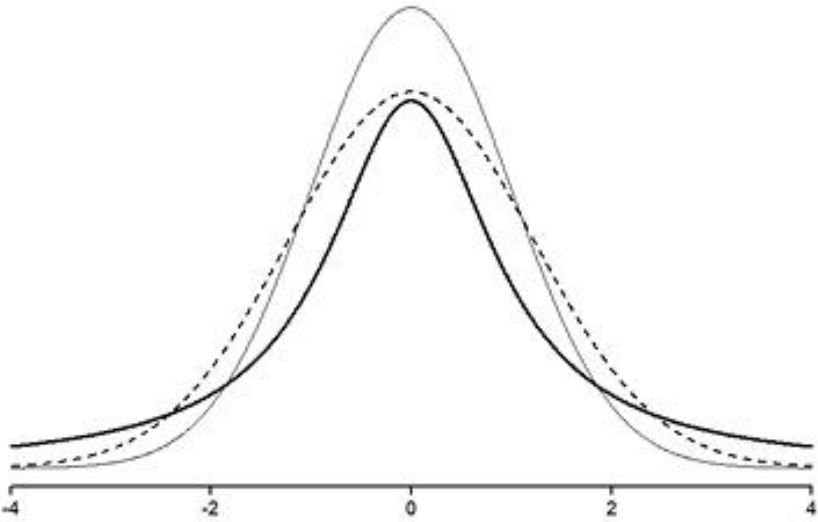
Figura 10.3



Ambas son simétricas, tienen forma de campana, pese a que una parece más “picuda” (menos “campanosa”) que la otra. Esto es delicadísimo, y logra engañar aun a estudiantes avanzados de probabilidad y estadística: la principal característica de la “normal” no es la simetría ni la forma de campana, sino la tasa a la que caen sus colas. Esto es difícil de tragar. Vamos por otro dibujito.

En la figura 10.4 les muestro dos densidades normales y también “la misteriosa”. La dibujada con la línea más gruesa es la “misteriosa”; las otras dos son normales, están centradas en el mismo lugar, con la salvedad de que una está más dispersa que la otra. Técnicamente, se dice que “tiene más varianza”, pero son normales ambas.

Figura 10.4



La enorme diferencia entre las normales y “la misteriosa” radica en que, hacia las puntas, las colas de la campana de Gauss caen más rápido en el caso de la normal, ¿lo ven?

Esto parece un tecnicismo para fans de Sheldon Cooper (el entrañable protagonista de The Big Bang Theory), pero tiene una consecuencia letal en términos de definir qué es raro y qué no lo es: cuando las colas caen lento, los eventos raros son mucho más probables de lo que parecen, tanto en términos absolutos como relativos. Dicho en criollo, cuando ciertos eventos son gobernados por la densidad misteriosa, los eventos raros no son tan raros como predice la normal.

Un tema central de los escritos de Taleb y de las finanzas, la política, la economía y otros eventos sociales es que creer tercamente en la distribución normal puede llevar a una grotesca subestimación de la ocurrencia de eventos raros.

Ahora sí, develemos el misterio. Un tipo de curva de densidad que se parece a lo que sugiere Taleb es la (bellísima y espantosa) densidad de Cauchy, llamada así por el notable matemático francés (Augustin-Louis, para más detalle). Una activa rama de las finanzas muestra que muchos fenómenos de esta disciplina (como los retornos de las acciones) parecen seguir este patrón "de colas pesadas", como se denomina al hecho de que la densidad de Cauchy caiga lento y habilite la (sobre)aparición de estos fenómenos raros.

Para peor, la ley de los grandes números, que ocupó un lugar prominente en el capítulo 1 y que habilita el uso de promedios para aprender probabilidades y otras magnitudes, deja de funcionar ante la presencia de datos que surgen de la distribución de Cauchy.

En definitiva, este tipo de fenómenos extremos es otra fuente de impredecibilidad que, para bien y para mal, nos acompaña desde los comienzos de la humanidad. De hecho, la mera existencia del universo tal vez entre en esta categoría.

Una cautela importante, y otro tema recurrente en este libro, es no llevar este argumento al extremo. Fenómenos como la designación del papa Francisco coexisten con la increíble habilidad de los algoritmos para predecir cuánto se tarda en ir en auto de una ciudad a la otra. En su afán por convencer a la sociedad de que no todo es predecible, a veces se cae en insistir con que nada lo es. Lo que hemos argumentado en este libro es que la realidad se mueve desde eventos simples, de fácil predecibilidad, hasta otros virtualmente inesperados (como los cisnes negros de Taleb). En el medio quedan sucesos que, si bien no se pueden predecir con exactitud, tienen algunas características predecibles, como los penales de la Selección (no se puede predecir con exactitud adónde van a parar, pero sí las chances de que un arquero ataje), algunas series económicas (para las cuales se puede construir una banda razonable para acotar sus movimientos) o los fenómenos del clima, en los que, en un plazo corto, se puede predecir con mucha precisión si va a llover. El argumento de Taleb es mostrar que el fenómeno de "colas pesadas" es otra de las fuentes de impredecibilidad, como las no linealidades de la meteorología, o la que surge de la naturaleza estratégica de los penales o el valor del dólar.

Donde coincidimos con el comunicador libanés es en la inocencia (consciente o no) de quienes racionalizan eventos raros, siempre con el diario de mañana, y que pretenden venderse como visionarios. Y también con que, por definición, no hay big data de eventos raros, es decir, es difícil encontrar el patrón detrás de eventos tales como “el proceso de designación de papas latinoamericanos”, porque, hasta la fecha, solo hubo uno, como cuando apareció el primer cisne negro en Australia. Negociar la tensión entre muchos datos irrelevantes versus pocos pero de gran utilidad, es un enorme desafío para la IA y la ciencia de la predicción moderna.

Volver del futuro

Este capítulo advierte acerca de los riesgos de (sobre)racionalizar fenómenos predictivos, o de abusar de algunas herramientas de pronóstico, como señala el problema de la “causalidad de Granger”. Y también respecto de que ciertos fenómenos son impredecibles por su naturaleza extrema y tal vez, en esas circunstancias, lo mejor sea aceptar su esencia imposible de pronosticar, sin perder tiempo en racionalizaciones inconducentes. En efecto, cuando los eventos son esencialmente impredecibles, “predecir el futuro es como tratar de conducir mirando el espejo retrovisor”, como dijo Peter Drucker.

La creación de historias compatibles con el pasado es un arma de doble filo. Por un lado, es el “test” al que ha apelado la ciencia para construir modelos y teorías que permiten adelantarse al futuro. Por el otro, muchas veces son meros ejercicios dialécticos para el autoengaño.

Es complejo evaluar la prevalencia del horóscopo. Que la ciencia muestre que no funciona y que a la vez se halle omnipresente en la sociedad indica que su existencia demanda una explicación no trivial.

La historia de Borges y Kafka muestra que la naturaleza predecesora de muchos eventos cruciales tiene un enorme componente de racionalización ex post, que se pueden contar “para atrás” en el tiempo, pero no necesariamente “para adelante”.

Una cosa es predecir un evento y otra, explicarlo. Muchas herramientas predictivas solo se adelantan al futuro, como los indicadores líderes.

Un error común es creer que, porque ha habido mejoras sustanciales en pronosticar un evento, se está más cerca de poder explicarlo.

La presencia de “cisnes negros” tiene que ver con sucesos que se encuentran en un rango extremo. No es simplemente “rareza”. Un cisne de 23.388 plumas es rarísimo, más allá de su color.

El fenómeno del cisne negro refiere a “raro y extremo”, es decir, lejano al resto de la ocurrencia de otros fenómenos comparables.

Si bien las finanzas, la política y la vida cotidiana contienen aspectos predecibles, están repletas de este tipo de eventos fundamentalmente impredecibles.

El problema práctico es que es complejo aprender patrones de eventos raros, justamente por falta de datos. Esto pone un límite obvio al sobreentusiasmo que muchos tienen con las posibilidades de big data y la IA.

Epílogo

Y hemos llegado al final de este recorrido por la ciencia de los pronósticos. Cansado, como quien regresa de un largo viaje en auto y solo pretende llegar a casa, le pedí a Roccatagliata que escribiera un epílogo para este libro. Obediente, lo hizo, pero puso frases como “que este libro sirva como un faro en la noche, guiándonos hacia un futuro lleno de posibilidades infinitas. Y que recordemos siempre que, en última instancia, somos los autores de nuestra propia historia”. Lo desecho, no sin antes recordarle que le pedí un epílogo y no un bolero.

Cual Seinfeld en el icónico episodio “The Betrayal” (en el que el tiempo transcurre hacia atrás: no se lo pierdan), déjenme resumir las principales ideas visitadas, pero en el orden estrictamente inverso al que las han encontrado en este libro.

Existen muchos aspectos de las sociedades que son esencialmente impredecibles, porque es como si estuviesen diseñados de modo explícito para serlo (como los penales en el fútbol). También varios fenómenos raros ocurren con más frecuencia de lo que uno sospecha, y es su naturaleza escasa lo que complica su análisis.

El problema es que, en efecto, algunos sucesos son impredecibles y, sin embargo, la historia de la ciencia dice que uno debería tener cuidado con adoptar posturas tan extremas. A modo de ejemplo, recordemos que antes de las teorías gravitatorias el cosmos se entendía como una conjunción de fenómenos relegados a lo fortuito y lo místico. Uno de los principales logros de la ciencia es, con una inteligente mezcla de observación y abstracción, encontrar orden y predecibilidad en sucesos en apariencia irracionalizables. Mientras escribo esto, acaba de suceder un eclipse solar en buena parte del hemisferio norte. El evento que llamó la atención no fue el eclipse, sino un mapa con datos de Airbnb que muestra que las tasas de alta ocupación en hoteles en los Estados Unidos siguen exactamente la

ruta del eclipse, evento complejo de mostrar hace no muchos años, e impensado hace cientos.

Muchos fenómenos de la economía y las finanzas son erráticos, pero, con las herramientas de series temporales y con mucha pericia, es posible captar y predecir algunas partes sistemáticas de su comportamiento, como lo que ocurre con las series estacionarias. Sobre este tema analizamos que la presencia de paseos al azar genera series virtualmente impredecibles, porque no son contenibles en ninguna banda estable. El origen de esta inestabilidad surge de acumular pasos aleatorios, como los del borracho o los precios de las acciones. Por desgracia, esta naturaleza hipervolátil de los paseos al azar es néctar y ambrosía para los aficionados a las predicciones con el diario de mañana, que son capaces de armar historias compatibles con cualquier pasado (porque el que busca encuentra).

Un importante fenómeno que complica al ámbito de las predicciones tiene que ver con las complejidades que surgen de la dinámica no lineal, como las que afectan a la meteorología y a eventos altamente interactivos, como el deporte profesional. Las simulaciones computacionales y la creación de escenarios alternativos cumplen (y cumplirán) un rol crucial para hacer predicciones en este tipo de eventos, amén de que hay mucho espacio para la IA y big data, como prometen las mejoras observadas en predecir el tiempo y la inflación.

La forma correcta de evaluar y diseñar herramientas predictivas es negociar la existencia inevitable de errores. La gran contribución de la IA-machine learning-big data es establecer una forma sistemática de lidiar con el compromiso entre sesgo y varianza: modelos predictivos demasiado complejos representan muy bien los datos del presente, pero predicen demasiado erráticamente los del futuro, y modelos demasiado simples son estables, pero sus pifies se confirman infaltables y sistemáticos. Se trata de encontrar un sano punto intermedio.

La sobreinterpretación de datos y patrones es tal vez el vicio que más afecta a la actividad predictiva. La dupla big data-IA puede ser tanto la solución como la enfermedad. Lo primero ocurrirá si la interacción entre datos y algoritmos puede corregir sesgos, para lo

cual el razonamiento abstracto (el que aportan la experiencia y la ciencia) es crucial. Lo segundo nos afectará en la medida en que creamos, inocentemente, que más datos es necesariamente mejor, o que LGN ocurre de forma gratuita y universal. Las fuertes dependencias entre los datos, la presencia de sesgos (aun ínfimos) y la recurrencia de eventos extremos ponen un límite natural a creer que la profusión de datos masivos resolverá cualquier problema informativo.

Las fallas de las ciencias sociales a la hora de predecir no se deben a la inoperancia de los analistas, sino a que en ellas confluyen, como en una pesadilla, todos los factores que atentan contra la predictibilidad: las fuertes heterogeneidades, las interacciones estratégicas y la relevancia de eventos extremos y raros.

La omnipresencia del azar es, tal vez, la razón por la que, paradójicamente, las predicciones funcionan peor cuando uno más las necesita. El azar es bravo, difícil de reconocer y complejo de interpretar y manipular. Separar cuánto hay de azar y cuánto de sistemático es, tal vez, la tarea más compleja de la predicción. Cuando todo es azar, los pronósticos son, en el mejor de los casos, limitados y, en el peor, completamente inefectivos. Cuando no hay azar, las predicciones son acertadas pero triviales.

Los pronósticos más relevantes son aquellos que refieren a eventos complejos y cruciales para la sociedad. El auténtico desafío de la actividad predictiva es alterar acciones, para la gestión del riesgo, para la toma de decisiones, para la evaluación de políticas. Lograr predicciones correctas no es simplemente haberle pegado o embocado a un resultado, sino demostrar que el pronóstico ex ante es confiable y útil, lo cual demanda una adecuada mezcla de experiencia y sabiduría, o la autoridad que se deriva de ambas.

Y así llegamos al punto de partida. Circunstancia interesante, llamativa: estoy en Tandil (¡creer o reventar!), para otra conferencia. Ya estoy en el auto, pronto a encarar la ruta. Mi hijo postadolescente me pregunta a qué hora pienso llegar, y no porque me extrañe, sino porque quiere saber cuándo se le acaba la libertad de disponer de la casa y la heladera. Roccatagliata y el Google Maps me dicen que en más o menos cuatro horas y media. Y me debato entre acelerar a fondo para enfrentarme cuanto antes al caos que

espero encontrar en el dulce hogar, o parar a comprar los exquisitos salamines tandilenses. Así que le respondo: “No lo sé”. Porque leí por ahí que la mejor manera de predecir el futuro es creándolo. Pero también pienso que la forma más lamentable de predecir es creando el futuro incorrecto.

Decálogo del buen pronosticador

El buen pronosticador…

…complementa promedios y probabilidades con otra información relevante sobre el tema, y luego hace sus predicciones. Por ejemplo, si un sistema de recomendación de series indica que una persona tiene el 0,5 de probabilidad de que le guste Friends, nuestro buen pronosticador eleva este número para personas mayores y lo baja para centennials.

…reconoce las cuatro fuentes de impredecibilidad: el azar (en casi todo), las interacciones estratégicas (como en los penales o las finanzas), la complejidad (como en la meteorología o los deportes) y la presencia no tan infrecuente de eventos extremos (como en la política o la cultura).

…sabe que una cosa es predecir si un evento ocurrirá y otra es corroborar las chances de que ocurra. En la mayoría de las circunstancias, lo primero es mucho más difícil que lo último. Quien dice “la probabilidad de que un candidato gane una elección es del 75%” está diciendo muchísimo más que “no tengo la menor idea”, aun cuando no pueda predecir con exactitud el resultado de esa elección.

…entiende que no hay pronósticos buenos o malos, sino pronosticadores buenos o malos. En efecto, una cosa es acertar un resultado y otra es hacerlo de manera sistemática, recurrente o con

algún argumento convincente. A primera vista, puede parecer una obviedad, pero predecir correctamente es convencer antes de que un resultado ocurra.

…percibe que muchos eventos de la vida cotidiana son esencialmente inciertos y no insiste en encontrar patrones y explicaciones luego de que esos fenómenos sucedieron. Es consciente de que las predicciones "con el diario de mañana" son inconducentes, a menos que se usen para realimentar algoritmos y corregir viejos errores de pronóstico.

…sabe que todos y cada uno de los pronósticos tienen un margen de error, y por eso mismo hace un uso inteligente de los datos para minimizarlo. Además, cuando es posible, lo mide y lo deja en claro en sus pronósticos.

…comprende que predecir mal es errar sistemáticamente o hacer pronósticos demasiado imprecisos.

…no cae en el error de pensar que big data implica necesariamente una gran cantidad de información. Esto supone que los datos que utilice deben cumplir con ciertas condiciones ideales, como las que rigen las encuestas profesionales o los experimentos científicos. Sin embargo, estos requisitos no siempre se cumplen: big data se caracteriza por cierta espontaneidad y anarquía.

…hace un uso inteligente de la tecnología. Es cauto con la utilización de los algoritmos y la inteligencia artificial (y es consciente de su potencial y sus limitaciones).

…jamás confunde predecir con explicar. Un algoritmo puede sacar provecho de la relación que hay entre lluvia y uso de paraguas para predecir el tiempo; pero esto no implica mejora alguna en la comprensión de por qué llueve, y menos aún de las maneras de hacer llover.

Apéndice

Fuentes, lecturas y sugerencias

He aquí una lista general de libros sobre pronósticos y temas relacionados, ordenados en dificultad técnica creciente. En la medida de lo posible, he intentado buscar versiones en castellano, pero esta materia es reciente y todavía no hay traducciones para varios de ellos.

Rojo, A. (2012), El azar en la vida cotidiana, Buenos Aires, Siglo XXI. Un muy lindo libro de divulgación sobre el rol del azar. Forma parte de esta misma colección.

Tetlock, P. E. y Gardner, D. (2016), Superpronosticadores. El arte y la ciencia de la predicción, Madrid, Katz. Este libro se focaliza en los predictores: quiénes pronostican correctamente, qué habilidades tienen.

Taleb, N. N. (2007), El cisne negro. El impacto de lo altamente improbable, Barcelona, Paidós. Un libro incisivo y polémico sobre la imposibilidad de lidiar con ciertos eventos extremos.

Siegel, E. (2013), Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die, Hoboken, Wiley. Un libro muy fácil de

leer sobre el uso de machine learning para actividades predictivas, con numerosos ejemplos y anécdotas, cero técnico.

Agrawal, A., Gans, J. y Goldfarb, A. (2018), Prediction Machines. The Simple Economics of Artificial Intelligence, Boston, Harvard Business Review Press [ed. cast.: Máquinas predictivas. La sencilla economía de la inteligencia artificial, Barcelona, Reverté, 2019]. Un lindo libro sobre qué significa pronosticar, qué valor les trae a las personas y a la sociedad, que focaliza en el rol de la IA.

Diebold, F. X. (2017), Forecasting in Economics, Business, Finance and Beyond, disponible en <n9.cl/n8uw5f>. Un libro técnico, para aquellos que han tomado una materia básica de estadística universitaria. Escrito por un autentico experto, es la introducción más simple al mundo de las series temporales.

James, G. y otros (2013), An Introduction to Statistical Learning. With Applications in R, Nueva York, Springer. De nivel técnico similar al anterior, es la introducción más simple y práctica a las técnicas más usuales de machine learning, con muchos ejemplos y código en R o Python. Este libro se puede descargar de forma libre y gratuita en <www.statlearning.com>.

Kuhn, M. y Johnson, K. (2013), Applied Predictive Modeling, Nueva York, Springer. De nivel un tanto más elevado que el anterior, este libro es un recorrido detallado por el universo de las técnicas de machine learning.

Elliott, G. y Timmermann, A. (2016), Economic Forecasting, Princeton, Princeton University Press. Un texto avanzado y actualizado sobre la visión moderna de series temporales para la predicción. Incorpora varias ideas del mundo de machine learning.

Murphy, K. P. (2012), Machine Learning. A Probabilistic Perspective, Cambridge (MA), MIT. Un texto avanzado y profundo sobre machine learning, que requiere una formación sólida en probabilidad y estadística.

Referencias y lecturas por capítulo

Capítulo 1

Dos hermosos libros sobre el azar son el de Alberto Rojo (ya mencionado) y el de Pablo Groisman (doy detalles más abajo). Uno más filosófico es del Jacovkis y Perazzo. La LGN ocupa un lugar central en el análisis de datos, la idea de "el teorema fundamental de la estadística" está tomada del libro de Davidson y MacKinnon. Sobre el clasificador de Bayes, una discusión muy interesante está en Hastie, Tibshirani y Freedman.

Davidson, R. y MacKinnon, J. G. (2004), Econometric Theory and Methods, Nueva York, Oxford University Press.

Groisman, P. (2023), Abrazar el azar, Buenos Aires, Eudeba.

Hastie, T., Tibshirani, R. y Friedman, J. (2009), The Elements of Statistical Learning. Data Mining, Inference, and Prediction, 2ª ed., Nueva York, Springer.

Jacovkis, P. y Perazzo, R. (2017), Azar, ciencia y sociedad, Buenos Aires, Eudeba.

Kasner, E. y Newman, J. (1985), Matemáticas e imaginación, Buenos Aires, Hyspamérica, “Biblioteca Personal de Jorge Luis Borges”.

Capítulo 2

Sobre encuestas sociales y medición de la pobreza y el bienestar, recomiendo el libro que escribimos con Leonardo Gasparini y Martín Cicowiez. El link de la Encuesta de Satisfacción Política y Opinión Pública que dirige Diego Reynoso en UdeSA está disponible en <n9.cl/atk74>. El paper de Meng (2018) es técnico, pero incluye un detallado análisis sobre el problema del impacto de pequeños sesgos en big data y cómo afectan a los pronósticos electorales, con un análisis de lo sucedido durante la elección de Trump en los Estados Unidos. Para una aplicación de estas ideas al monitoreo del covid-19, Bradley y otros (2021). Zhou y otros (2021) discurren sobre cómo los datos de redes sociales se pueden usar para mejorar la performance de las encuestas estándar. Un libro clásico sobre muestreo (y accesible para quienes hayan tomado una materia de estadística) es el de Lemeshow y Levy (1999).

Bradley, V.C., Kuriwaki, S., Isakov, M. y otros (2021), “Unrepresentative Big Surveys Significantly Overestimated US Vaccine Uptake”, Nature, 600: 695-700.

Gasparini, L., Cicowiez, M. y Sosa Escudero, W. (2013), Pobreza y desigualdad en América Latina. Conceptos y herramientas

analíticas, Buenos Aires, Temas.

Lemeshow, S.y Levy, P. (1999), Sampling of Populations. Methods and Applications, 3ª ed., Nueva York, Wiley.

Mayer-Schönberger, V. y Cukier, K. (2013), Big Data. A Revolution that Will Transform How We Live, Work, and Think, Boston, Houghton Mifflin Harcourt.

Meng, X.-L. (2018), “Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 Us Presidential Election”, The Annals of Applied Statistics, 12(2): 685-726.

Zhou, Z. y otros (2021), “Why Polls Fail to Predict Elections”, Journal of Big Data, 8, 137.

Capítulo 3

La historia de la estadística, el problema de regresión a la media y un relato sobre Galton se halla en Stigler. El relato del militar y los soldados está en Kahnemann (2012). Una jugosa discusión (técnica) sobre el problema de regresión a la media es Koenker (2016). Un análisis detallado y moderno del problema de tanques alemanes figura en Clark y otros (2021).

Clark, G., Gonye, A. y Miller, S. J. (2021), “Lessons from the German Tank Problem”, arXiv:2101.08162.

Hotelling, H. (1933), “Review: The Triumph of Mediocrity in Business by Horace Secrist”, Journal of the American Statistical Association, 28(184): 463-465.

Kahnemann, D. (2012), Pensar rápido, pensar despacio, Buenos Aires, Debate.

Koenker, R. (2016), “The Regression Fallacy”, mimeo, University of Illinois Urbana-Champaign.

La Nación (2019), “¿Una maldición? Qué pasó con los actores que ganaron el Martín Fierro Revelación”, 6 de junio (retomado en el texto de este capítulo).

Secrist, H. (1933), The Triumph of Mediocrity in Business, Chicago, Bureau of Business Research, Northwestern University.

Stigler, Stephen M. (1986), The History of Statistics. The Measurement of Uncertainty Before 1900, Cambridge (MA), The Belknap Press of Harvard University.

Capítulo 4

El paper de Diaconis sobre lanzamiento de monedas es técnico, pero la idea principal se entiende perfectamente. El problema de Monty Hall es un clásico de las probabilidades, pero sigue dándole dolores

de cabeza a cualquiera que lo enfrente y ha sido estudiado hasta el hartazgo. Un análisis reciente se halla en Alvarado (2024). Si quieren jugar online y ver si les conviene cambiar de puerta, un simulador está disponible en <n9.cl/58wyu>. La discusión más simple (y con muchísima autoridad) sobre probabilidades, promedios, variables aleatorias y probabilidades condicionales consta en Friedman, Pisani y Purves, y no requiere más matemática que la del secundario. Sobre pronósticos y pronosticadores, Superforecasting es una gran lectura. La entrada de "German tank problem" (y su versión castellana "Problema de los tanques alemanes") de Wikipedia tiene una derivación detallada de la fórmula. Una explicación clara, con una simulación, puede verse en Johnson (1994). La explicación de Steve Pinker está en <n9.cl/zm1rr>.

Alvarado, R. C. (2024), "What's So Hard about the Monty Hall Problem?", arXiv, disponible en <n9.cl/ly92i>.

Diaconis, P., Holmes, S. y Montgomery, R. (2007), "Dynamical Bias in the Coin Toss", SIAM Review, 49(2): 211-235.

Freedman, D., Pisani, R. y Purves, R. (2007), Statistics, 4ª ed., Nueva York, W. W. Norton & Co.

Johnson, R. W. (1994), "Estimating the Size of a Population", Teaching Statistics, 16(2): 50-52.

Tetlock, P. E. y Gardner, D. (2015), Superpronosticadores, ya citado.

Capítulo 5

Cualquier libro de machine learning tiene una discusión seria sobre sesgo, varianza y error cuadrático medio; el de James y otros (2013), ya mencionado, es altamente recomendable. Tuve la enorme fortuna de estudiar con Roger Koenker, uno de los padres del tratamiento técnico de errores asimétricos; su monografía de 2005 es la biblia sobre esta cuestión. La investigadora argentina Luciana Ferrer, de la Facultad de Ciencias Exactas y Naturales de la UBA, es toda una autoridad en los problemas que atan la predicción de probabilidades con la de eventos. El texto de Murphy es una referencia avanzada sobre todos estos temas. Acerca del método de Montecarlo y sus aplicaciones, la referencia es el libro de Robert y Casella (si bien muy avanzado). El ejercicio del globo terráqueo está tomado de libro de Richard McElreath, que promete revolucionar la enseñanza de la estadística en tiempos de machine learning y computadoras.

James, G. y otros (2013), An Introduction to Statistical Learning (también citado en la lista inicial).

Koenker, R. (2005), Quantile Regression, Cambridge, Cambridge University Press.

McElreath, R. (2020), Statistical Rethinking. A Bayesian Course with Examples in R and Stan, 2ª ed., Boca Ratón, CRC Press.

Murphy, K. P. (2012), Machine Learning, ya citado.

Robert, C. P. y Casella, G. (2004), Monte Carlo Statistical Methods,

2ª ed., Nueva York, Springer.

Capítulo 6

El libro de Mindlin, en esta colección, es una lectura simple sobre caos y complejidad. Sobre la historia de la meteorología, recomiendo la extensa monografía de Benjamin y otros. Un fascinante libro sobre los orígenes de los pronósticos del clima es el de Peter Moore. El paper sobre el potencial de la IA en los pronósticos del tiempo es de Remi Lam y otros. El sitio del proyecto de predicciones del Mundial en la Facultad de Ciencias Exactas y Naturales de la UBA está en <301060.exactas.uba.ar>. El periodista argentino Marcelo Gantman es todo un experto en cuestiones de datos, algoritmos y deporte. Pueden encontrar sus artículos y su muy recomendable libro en <bigdatasports.media>.

Benjamin, S. G. y otros (2018), “100 Years of Progress in Forecasting and NWP Applications”, Meteorological Monographs, 59(1): 13.1-13.67.

Gantman, M. (2022), El final del deporte como lo conocimos, disponible en <www.elfinaldeldeporte.com.ar>.

Kalnay, E. (2003), Atmospheric Modeling, Data Assimilation and Predictability, Cambridge, Cambridge University Press.

Lam, R. y otros (2023), “Learning Skillful Medium-Range Global Weather Forecasting”, Science, 382(6677): 1416-1421.

Mindlin, G. (2008), Causas y azares. La historia del caos, Buenos Aires, Siglo XXI, col. “Ciencia que Ladra”.

Moore, P. (2016), The Weather Experiment. The Pioneers who Sought to See the Future, Nueva York, Farrar, Straus and Giroux.

Capítulo 7

Mi libro de 2019 se propone como divulgación sobre las herramientas de machine learning y el mundo de big data. Mi otro libro, de 2022, es una introducción simple a varias ideas de pronósticos, en particular de los métodos de regresión. La introducción técnicamente más simple al mundo de machine learning es James y otros (2013), que ya asoma por tercera vez en esta bibliografía y está destinado a convertirse en un clásico. En orden creciente de dificultad, los textos de Bishop y Murphy son muy recomendables, si bien matemáticamente demandantes. El de Goodfellow y otros es “el” libro de deep learning. Lamentablemente, la literatura sobre este tema es muy “polar”: o es demasiado simplista o muy técnica. El libro de Goodfellow pertenece a este grupo, pero está escrito con muchísima autoridad. Una excelente explicación introductoria de deep learning y su relación con ChatGPT es Wolfram (2023).

Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Nueva York, Springer.

Goodfellow, I., Bengio, Y. y Courville, A. (2016), Deep Learning, Cambridge (MA), MIT Press.

Murphy, K. P. (2012), Machine Learning, ya citado.

Sosa Escudero, W. (2019), Big data, 9ª ed., Buenos Aires, Siglo XXI, col. “Ciencia que Ladra”.

— (2022), Qué es (y qué no es) la estadística, 2ª ed. rev., Buenos Aires, Siglo XXI, col. “Ciencia que Ladra”.

Wolfram, S. (2023), What is ChatGPT Doing… And Why Does It Work?, disponible en <n9.cl/6jbdu>.

Capítulo 8

El universo de los pronósticos con series temporales es inmenso. El texto más accesible es el de Diebold (2017) (repleto de ejemplos y casos), que se puede bajar gratuitamente de la página del autor. Una referencia clásica e ineludible (y avanzada) es Hamilton (1994). La referencia moderna sobre pronósticos económicos es el libro de Elliott y Timmerman. Wei (2006) contiene muchas aplicaciones en varios ámbitos, si bien es un texto avanzado. Sobre la relevancia de los “paseos al azar”, el libro de Malkiel es de muy fácil lectura, sin ningún tecnicismo. Su némesis es el de Lo y MacKinlay (2001), aunque muy técnico. Lo y MacKinlay son también autores de un libro muy recomendable de econometría financiera avanzada, escrito con mucha autoridad. Sobre el potencial de ChatGPT para pronosticar la inflación, recomiendo el estudio reciente de Faria e Castro y Lebovici. Hay un lindo simulador online de random walks en <n9.cl/0dpld>.

Diebold, F. X. (2017), Forecasting in Economics, Business, Finance

and Beyond, ya citado.

Elliott, G.y y Timmermann, A. (eds., 2016), Economic Forecasting, también citado.

Faria e Castro, M. y Lebovici, M. (2023), “Artificial Intelligence and Inflation Forecasts”, Federal Reserve Bank of St Louis, Working Paper 2023-015D.

Hamilton, J. D. (1994), Time Series Analysis, Princeton, Princeton University Press.

Lo, A. W. y MacKinlay, A. C. (1997), The Econometrics of Financial Markets, Princeton, Princeton University Press.

— (2001), A Non-Random Walk down Wall Street, Princeton, Princeton University Press.

Malkiel, B. G. (2019), A Random Walk down Wall Street. The Time-Tested Strategy for Successful Investing, 12ª ed., Nueva York, W. W. Norton & Co [ed. cast.: Un paseo aleatorio por Wall Street. La estrategia para invertir con éxito, Madrid, Alianza, 2016].

Wei, W. W. S. (2006), Time Series Analysis. Univariate and Multivariate Methods, 2ª ed., Boston, Pearson Addison Wesley.

Capítulo 9

Un interesantísimo libro sobre estrategias y juegos es el de Dixit y Nalebuff, muy accesible, no requiere ninguna formación técnica. Un texto simple de teoría de juegos es el de Gibbons. Una obra avanzada (sobre la cual he basado varias de las ideas de este capítulo) es el de Osborne y Rubinstein. Sobre expectativa y el "principio de incertidumbre" en economía, propongo la lectura del artículo de David Levine, un enorme de la teoría microeconómica. Acerca de penales, fútbol y estrategia, la máxima autoridad mundial es el investigador español Ignacio Palacios Huerta. El libro de Taleb es una excelente referencia sobre las enormes dificultades que conlleva lidiar con el azar. El artículo de Adrián Paenza sobre el azar y la dificultad de "producir el azar" es una gran puerta de entrada al tema.

Dixit, A. K. y Nalebuff, B. J. (2008). The Art of Strategy. A Game Theorist's Guide to Success in Business and Life, Nueva York, W. W. Norton & Co [ed. cast.: El arte de la estrategia. La teoría de juegos, guía del éxito en sus negocios y en su vida diaria, Barcelona, Antoni Bosch, 2010].

Gibbons, R. (1992), Game Theory, Princeton, Princeton University Press [ed. cast.: Un primer curso de teoría de juegos, Barcelona, Antoni Bosch, 2022].

Levine, D. (2012), "Why Economists Are Right: Rational Expectations and the Uncertainty Principle in Economics", HuffPost, disponible en <n9.cl/royfes>.

Osborne, M. J. y Rubinstein, A. (1994), A Course in Game Theory,

Cambridge (MA), MIT.

Paenza, A. (2010), “Azar”, Página/12, 18 de julio.

Palacios Huerta, I., (2003), “Professionals Play Minimax”, The Review of Economic Studies, 70(2): 395-415.

— (2023), “Maradona Plays Minimax”, Sports Economics Review, 1.

Taleb, N. N. (2008), Fooled by Randomness. The Hidden Role of Chance in Life and in the Markets, 2ª ed., Nueva York, Random House [ed. cast.: ¿Existe la suerte? Las trampas del azar, Barcelona, Paidós, 2009].

Capítulo 10

Recomiendo enfáticamente leer el ensayo de Borges sobre Kafka, es de una precisión quirúrgica y una profundidad admirable. En mi libro de 2020 me ocupé extensamente del nexo entre el mundo de los datos y los algoritmos, y el del universo del gran autor argentino. El trabajo sobre horóscopo y pronósticos es de Carlson. Pardo tiene una discusión moderna del “efecto Barnum” y la psicología del horóscopo. Sobre el potencial de big data y la IA para monitorear y pronosticar varios aspectos de la economía y la sociedad, recomiendo los artículos de Raj Chetty, firme candidato al Nobel. El trabajo seminal de causalidad de Granger se publicó en 1969. Hay un fascinante relato sobre la distribución de Cauchy en el recomendable libro de Stigler. El libro de Taleb es largo, idiosincrático, polémico, autorreferente, pero inevitable.

Borges, J. L. (1952 [1951]), “Kafka y sus precursores”, en Otras inquisiciones, Buenos Aires, Sur.

Carlson, S. (1985), “A Double-Blind Test of Astrology”, Nature, 318: 419-425.

Chetty, R. y otros (2022), “Social capital I: Measurement and Associations with Economic Mobility”, Nature, 608(7921): 108-121.

— (2022), “Social Capital II: Determinants of Economic Connectedness”, Nature, 608(7921): 122-134.

— (2024), “The Economic Impacts of Covid-19: Evidence from a New Public Database Built Using Private Sector Data”, The Quarterly Journal of Economics, 139(2): 829-889.

Currey, R. (2011), “U-Turn in Carlson Astrology Test?”, Correlation, 27(2): 7-33.

Granger, C. W. J. (1969), “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods”, Econometrica, 37(3): 424-438.

Pardo, E. (2022), “The Psychology behind Why We Believe in Horoscopes”, disponible en <n9.cl/rhcdv7>.

Sosa Escudero, W. (2020), Borges, big data y yo, Buenos Aires, Siglo XXI, col. “Ciencia que Ladra”.

Stigler, S. (1999), “Cauchy and the Witch of Agnesi”, en Statistics on the Table, Cambridge, Harvard University Press.

Taleb, N. N. (2007), El cisne negro, ya citado.

Thurman, W. N. y Fisher, M. E. (1988), “Chickens, Eggs and Causality, or Which Came First?”, American Journal of Agricultural Economics, 70(2): 237-238.

Por último, comparto un QR al sitio:

<waltersosaescudero.weebly.com>

Ahí encontrarán lecturas, ejemplos, ejercicios y código computacional, incluidos los que usé para ilustrar las historias de este libro.

Agradecimientos

El libro que tienen en sus manos es una versión muy mejorada del manuscrito que entregué en su momento a la editorial. Raquel San Martín (editora de Siglo XXI) y Diego Golombek (director de Ciencia que Ladra) leyeron, comentaron y discutieron en detalle este trabajo y contribuyeron sustancialmente a mejorarlo. Marisa García Fernández y Luciano Padilla López aportaron su reconocido profesionalismo y su calidez en la etapa final. De Carlos Díaz (director de Siglo XXI) recibí su habitual entusiasmo con mis proyectos de divulgación científica.

Y si de divulgación hablábamos, comento que este es un trabajo un tanto más "coral" que mis anteriores libros. Si bien el hilo conductor es el de mi especialización (econometría, estadística, ciencia de datos), el proyecto que aquí les ofrezco implicó meterme en la cocina de varias disciplinas. Tuve la enorme fortuna de toparme con talentosos "chefs" que, con generosidad, me compartieron sus secretos. Por eso agradezco a Pablo Groisman (probabilidades y azar), Diego Reynoso y Federico Tiberti (encuestas electorales), Javier García Cicco (pronósticos macroeconómicos y financieros), Luciana Ferrer (predicción y probabilidades), Willy Durán y Marcelo Gantman (deportes), Guadalupe Nogués (complejidad) y María Eugenia Dillon (meteorología). Daniela Parada aportó sus experiencias docentes con el material de mis libros. Cualquier mérito que tenga este actual es en gran medida de ellos (no así los yerros y omisiones, que son todos de elaboración propia).

Buena parte de estas páginas fue escrita durante la licencia sabática que pasé en el Instituto de Cálculo de la Facultad de Ciencias Exactas de la Universidad de Buenos Aires. Agradezco a todos sus miembros por la cálida hospitalidad y el ambiente motivador. Como le dije una vez al decano, Willy Durán, el Instituto es una suerte de Disneylandia para curiosos y entusiastas de la ciencia y el

conocimiento, una verdadera isla de excelencia de nivel internacional. Como siempre, la Universidad de San Andrés (UdeSA, mi afiliación principal) proveyó los insumos fundamentales para pensar y desarrollar este proyecto: tiempo libre, generosos recursos logísticos e intelectuales, un ámbito desafiante y un apoyo incondicional a mi tarea de divulgación.

Una vez más, Sebastián Campanario me brindó su espacio en "Álter Eco" –su columna en La Nación, donde aparecieron algunas historias que luego se incorporarían aquí–, además de su habitual entusiasmo y creatividad para con la difusión de nuevas ideas. Emiliano Chamorro (Instituto Baikal), Daniel Gonzalez Isolio, Daniel Serrot y Ariel Urcola (UdeSA) me invitaron a dar cursos y charlas sobre la temática que capítulo a capítulo seguí charlando con ustedes.

Daniela Rodríguez, Guillermo Solovey, Ana Bianco, María Edo, Mariana Marchionni, Ricardo Bebczuk, Leonardo Gasparini, Javier Alejo, Mercedes Iacoviello, Martín Tetaz, Christian Ruzzier, Lucía Quesada, María Gabriela Ertola Navajas y Lorena Garegnani aportaron lecturas, referencias, contactos, datos, ideas, comentarios o, simplemente, palabras de apoyo enfático para este proyecto. Lorenzo T Roccatagliata aportó su reconocida eficiencia en varias tareas operativas. Agradezco especialmente a mis profesores, Hildegart Ahumada, Chung-ming Kuan, Roger Koenker y Anil Bera, con quienes me formé en econometría y estadística.

Como siempre, mi esposa Mercedes, mi hijo Alejandro y mi madre Mary soportaron estoicamente la histeria de la etapa final de la elaboración de este libro.